Essays in Labor Economics and Contract Theory

Thesis advisor Author
Professor Lawrence Katz **Neel Dattatray Rao**

Essays in Labor Economics and Contract Theory

Abstract

This dissertation consists of three essays in labor economics and contract theory.

The first essay examines whether one's wage is based on information about the performance of one's personal contacts. I study wage determination under two assumptions about belief formation: individual learning, under which employers observe only one's own characteristics, and social learning, under which employers also observe those of one's personal contacts. Using data on siblings in the NLSY79, I test whether a sibling's characteristics are priced into one's wage. If learning is social, then an older sibling's test score should typically have a larger adjusted impact on a younger sibling's log wage than vice versa. The empirical findings support this prediction. Furthermore, I perform several exercises to rule out other potential factors, such as asymmetric skill formation, human capital transfers, and role model effects.

The second essay analyzes the influence of macroeconomic conditions during childhood on the labor market performance of adults. Based on Census data, I document the relationship of unemployment rates in childhood to schooling, employment, and income as an adult. In addition, a sample from the PSID is used to study how the background attributes of parents raising children vary over the business cycle. Finally, information from the NLSY79-CH is examined in order to characterize the impact of economic fluctuations on parental caregiving. Overall, the evidence is consistent with a negative effect of the average unemployment rate in childhood on parental investments in children and the stock of human capital in adulthood.

The third essay studies the bilateral trade of divisible goods in the presence of stochastic transaction costs. The first-best solution requires each agent to transfer all of her good to the other agent when the transaction cost reaches a certain threshold value. However, in the absence of court-enforceable contracts, such a policy is not incentive compatible. We solve for the unique maximal symmetric subgame-perfect equilibrium, in which agents can realize some gains from trade by transferring their goods sequentially. Several comparative statics are derived. In some cases, the first-best outcome can be approximated as the agents become infinitely patient.

Contents

Acknowledgments

I am thankful to Lawrence Katz, Raj Chetty, and Oliver Hart for their willingness to advise this dissertation and for their guidance during the writing process.

The chapters in this dissertation also benefitted from the comments of several individuals. Roland Fryer, Edward Glaeser, and seminar participants at Harvard University provided useful feedback on Chapter 1. A discussion with Rajeev Dehejia as well as comments by seminar participants at Harvard University helped improve Chapter 2. Drew Fudenberg, Michihiro Kandori, Takuo Sugaya, and Yasutora Watanabe made valuable suggestions on Chapter 3.

I am especially grateful to Yuichiro Kamada with whom I coauthored the third chapter of this dissertation.

Chapter 1

Social Learning in the Labor Market: An Analysis of Siblings

1.1 Introduction

An important question in labor economics concerns the mechanisms through which personal contacts influence job search behavior and wage setting decisions. As Granovetter's (1974) classic survey of workers in the Boston area illustrates, approximately half of all jobs are obtained with the help of a social contact. The extensive use of friends, relatives, and acquaintances in job search makes it possible for personal contacts to play a role in shaping employers' beliefs about a worker's skills. As Rees (1966) notes when studying workers in a Chicago neighborhood, "Present employees tend to refer people like themselves, and they may feel that their own reputation is affected by the quality of the referrals." Likewise, Montgomery (1991) presents a model of job search through personal contacts in which workers belonging to the same reference group are endowed with similar abilities and firms make wage offers to referred workers based on the performance of current employees. Given the importance of informal social ties in finding a job, an analysis of the effects of social networks on the wage structure appears to be essential for a complete understanding of the functioning of labor markets.

This paper develops and implements an empirical test for whether a worker's wage incorporates information on the performance of her personal contacts. Combining a sibling model similar to Griliches (1979) with a learning model related to Altonji and Pierret (2001), I construct a framework in which workers are organized into disjoint social groups composed of a small number of agents with correlated abilities and differing ages, and I examine wage determination under two competing assumptions about the market's formation of beliefs: individual learning and social learning. Under individual learning, a worker's wage is set equal to the conditional expectation of her productivity given only her own schooling and performance, whereas under social learning, a worker's wage is set equal to the conditional expectation of her own productivity given the schooling and performance of all the members of her social group, including herself.

Using sibling data from the NLSY79, I apply this framework to test for a form of statistical nepotism in which a worker's wage can be decomposed into a component based on a sibling's performance as well as a component based on one's own performance. The basic logic behind this test is as follows. If one sibling is older than another sibling, then employers should have more precise information about the older sibling, because the older sibling's performance is likely to have been observed for a longer length of time. Consequently, when market participants form Bayesian beliefs about the abilities of the two siblings, the older sibling's average performance would have a greater impact on employers' mean beliefs about the younger sibling's ability than vice versa. Assuming that the labor market is competitive, this asymmetry should be reflected in the wages of the two siblings. Hence, the component of the younger sibling's wage attributable to the older sibling's ability would be larger than the component of the older sibling's wage attributable to the younger sibling's ability.

Empirically, given data on the test scores and schooling of siblings, this weighting can be detected by regressing an individual's log wage on her own and a sibling's test scores and schooling. As in much of the literature on employer learning, the test scores in the data are treated as being known to the econometrician but not directly observable to employers. If employer learning is nepotistic in nature, then the ratio of the coefficient on a sibling's test score to the coefficient on one's own test score should typically be higher in a younger sibling's log wage than in an older sibling's log wage. However, if employer learning is entirely individual, then the ratio of the coefficient on a sibling's test score to the coefficient on one's own test score should be the same for both a younger and an older sibling. In addition to performing this simple test, I document several pieces of evidence indicating that the main patterns observed in the data are unlikely to be explained by factors unrelated to the learning processes studied in this paper.

The empirical strategy here integrates elements from five largely distinct literatures in labor economics. First, this paper is part of a sizeable literature on the identification of social effects.[1] The most closely related paper in this literature is Case and Katz (1991), which attempts to detect neighborhood influences by regressing an individual's outcome variable on the background variables of her peers. The current paper tests for social learning by regressing a worker's log wage on a sibling's test score as well as other control variables. In addition, I seek to address the concerns of Manski (1993) regarding the difficulties in distinguishing between social and nonsocial effects by focusing on the relative values of the coefficients on an older and a younger sibling's test scores instead of the absolute value of the coefficient on a sibling's test score in itself. Furthermore, because siblings form a clearly defined social unit, the use of sibling data mitigates some of the econometric problems associated with the misspecification of peer groups. By contrast, when using information on friends, such issues may be more severe.

Second, this paper belongs to a long line of research that exploits the special structure of sibling data to address a variety of questions in labor economics. As noted above, the model of social groups used in this paper is based on the sibling model in Griliches (1979). Moreover, sibling data appears to be relatively well suited for the purpose of analyzing social effects in employer learning, because non-twin siblings tend to have a moderately high correlation in ability. By contrast, if individuals were assigned to social groups mostly at random as in some quasi-experimental designs, then an individual's performance might provide little information from which employers could infer the ability of her peers, and if individuals in the same social group had very similar characteristics as could be the case with identical twins, then it might be difficult to distinguish empirically between the components of a person's wage based on her own and her peer's performance.

Third, this paper contributes to a growing literature on employer learning. In order to examine social interactions in the employer learning process, I extend the basic methodology developed by Farber and Gibbons (1996) and Altonji and Pierret (2001). Given the assumption that the AFQT scores in the NLSY79 are not directly observable to employers, Altonji and Pierret (2001) develop a test for statistical discrimination, in which employers use a worker's easily observable characteristics to infer her productive ability. Those authors find that employers statistically discriminate on the basis of education but not race. The current paper devises a test for statistical nepotism, in which employers infer an individual's productivity based partly on information about her relatives.

[1] See Ioannides and Loury (2004) for a review of existing research on social effects in labor markets.

Fourth, this paper is relevant to a theoretical literature on social networks in labor markets. The framework in the current paper is most similar to the model in Montgomery (1991), which examines job search through homophilous social networks. In that model, workers are arranged into social groups containing either one or two members, and social groups of size two consist of an older and a younger worker with correlated abilities. Because employers initially have imperfect information about each worker's ability, they use the observed performance of the more senior worker in each pair to infer the ability of her more junior counterpart. The empirical analysis in the current paper can be regarded as a test of the basic information transmission mechanism in Montgomery (1991), whereby employers learn about a younger peer's productivity from her older peer's performance.

Fifth, this paper contributes to a small empirical literature that attempts to test for nepotism in a variety of labor market settings. The most closely related paper in this literature is Lam and Schoeni (1993), whose empirical strategy involves comparing the coefficient on a father's schooling to the coefficient on a father-in-law's schooling in a wage equation.[2] Using data from Brazil, those authors find that a father-in-law's schooling has a stronger impact on a worker's wage than a father's schooling and interpret this result as evidence against the hypothesis that the positive coefficient on parental education in wage equations is due to nepotistic connections. In addition, there is at least some existing evidence suggesting that siblings may have an important impact on labor market outcomes. Rees and Gray (1982) observe that the presence of an employed older sibling is associated with a significantly higher probability of being employed. Those authors find no evidence that parents affect a youth's employment outcomes.

The main empirical results in this paper are consistent with the presence of statistical nepotism in the labor market. That is, I find that an older sibling's test score has a significantly larger impact on a younger sibling's log wage than a younger sibling's test score has on an older sibling's log wage. Nonetheless, there are mechanisms other than social learning that could generate similar patterns in the data, including asymmetries in skill formation, human capital transfers between siblings, and role model effects. Therefore, I perform a number of additional exercises to differentiate social learning from other possible explanations for the results. First, I provide evidence of

[2]Similarly, Hellerstein and Morrill (2011) examine trends in the transmission of human capital from fathers to daughters by analyzing changes in the likelihood that a woman enters her father's as compared to her father-in-law's occupation.

differences in job search behavior among siblings of different ages, documenting a robust positive correlation between birth order and the probability of obtaining one's job with the help of a sibling. This finding makes it plausible that a sibling might exert a direct influence on labor market outcomes with older siblings having a larger effect on younger siblings than vice versa. Second, I show that there is no evidence that a person's test score is more strongly related to an older sibling's test score than to a younger sibling's test score. Third, I also find no evidence that an older sibling's test score has a greater impact on a younger sibling's schooling than a younger sibling's test score has on an older sibling's schooling. These two results cast doubt on the role of asymmetric skill formation, human capital transfers, and role model effects in explaining the findings regarding the log wage. Fourth, while an older sibling's test score is seen to have a greater impact on a younger sibling's log wage than vice versa, the same regressions indicate that an older sibling's schooling has a smaller impact on a younger sibling's log wage than vice versa. This result suggests that an informational mechanism like employer learning may be helpful in explaining the asymmetries observed in the log wage regressions, because a hard-to-observe ability measure like test scores is seen to exhibit a different pattern of behavior than an easy-to-observe ability measure like schooling levels.

In order to provide an additional layer of evidence that labor market interactions in general and social learning in particular are responsible for generating the sibling effects observed in the log wage regressions, I report the results of three falsification tests. First, I demonstrate that the test scores of older and younger siblings without substantial labor market experience have similar impacts on a person's log wage. Second, I find no discernable difference between the impacts of the test scores of older and younger siblings who were living in different geographic regions during the early stages of their careers. Third, there is some evidence of a decrease in the impact of a sibling's test score when two siblings initially residing in the same region become geographically separated. Overall, the findings from these three tests, as well as the four facts listed above, seem to isolate social learning in the labor market as the most compelling explanation for the main empirical results.

The remainder of this paper is structured as follows. Section 1.2 presents the basic learning models that serve as a framework for the empirical investigation in this paper. This section contains four parts, the first of which develops a statistical model of the labor market characteristics of siblings, the second of which examines the properties of schooling and test scores as ability measures, the third of which studies the structure of log wages under individual learning, and the fourth of which explores wage determination under social learning. Section 1.3 outlines several extensions of the basic learning models in the previous section. Section 1.4 addresses some estimation issues that arise because the regression coefficients are predicted to change with one's own age and possibly those of one's siblings. Section 1.5 discusses the construction of the main estimation sample. Section 1.6 presents the empirical results. This section contains five parts, the first of which studies job search behavior among siblings, the second of which analyzes the relationships among the test scores, schooling levels, and log wages of siblings, the third of which performs various robustness checks on the main estimation results, the fourth of which tests some further implications of the learning models, and the fifth of which discusses the three falsification exercises mentioned above. Section 1.7 summarizes the findings of this paper and concludes.

1.2 Sibling Models with Employer Learning

This section discusses the empirical implications of employer learning for the observed relationship between siblings' test scores and log wages. Section 1.2.1 presents a model of the labor market characteristics of a pair of siblings. The framework developed in this section embeds a sibling model based on Griliches (1977, 1979) into a learning model related to Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007). Section 1.2.2 characterizes the statistical relationships among siblings' test scores, schooling choices, and ability levels. The results derived there shed light on the determinants of sibling correlations in human capital measures and facilitate the analysis of employer learning in the subsequent sections. Section 1.2.3 examines the relationship between test scores and log wages if employers use information only on an individual's own schooling and performance when setting wages, and section 1.2.4 explores how this relationship changes if employers also use information on a sibling's schooling and performance.

1.2.1 Labor Market Characteristics of Siblings

This section presents a statistical model of siblings' labor market attributes. The treatment here focuses on the case in which there are two siblings, 1 and 2.[3] As in much of the literature on

[3]See appendix A.4 for an extension of the model to include an arbitrary number of siblings.

employer learning,[4] the log labor productivity $l(s_i, a_i, t_i)$ of person $i \in \{1, 2\}$ is assumed to be decomposable into two components:

$$l(s_i, a_i, t_i) = g(s_i, a_i) + h(t_i), \tag{1.1}$$

where $g(s_i, a_i)$ is a time-invariant component of productivity, and $h(t_i)$ represents additional human capital accumulated with age t_i. Letting $\beta > 0$, the function $g(s_i, a_i)$ is linear in schooling s_i and ability a_i:

$$g(s_i, a_i) = \beta s_i + a_i, \tag{1.2}$$

where the coefficient on a_i is without loss of generality normalized to one. Note that the production function in equation (1.2) does not include any direct interactions between siblings.

The abilities a_1, a_2 of the two siblings are joint normally distributed with respective means μ_{a1} and μ_{a2}, identical variance $\sigma_a^2 > 0$, and correlation $\rho_a \in (0, 1)$. Letting $\gamma > 0$, schooling is related to ability through:

$$s_i = \gamma a_i + \epsilon_i, \tag{1.3}$$

where ϵ_i, which represents factors other than labor market ability that influence education decisions, is assumed to be independent of a_1, a_2. The error terms ϵ_1, ϵ_2 are joint normally distributed with respective means $\mu_{\epsilon 1}$ and $\mu_{\epsilon 2}$, identical variance $\sigma_\epsilon^2 > 0$, and correlation $\rho_\epsilon \in (0, 1)$. Specification (1.3) is consistent with a setting in which siblings do not affect each other prior to entry into the labor market and choose schooling levels independently based on their own ability.[5]

The information available to employers about ability a_i is symmetric but imperfect. In particular, employers observe the schooling s_i of each person as well as a sequence $r_i = \{r_{iu}\}_{u=1}^{t_i}$ of noisy productivity signals given by:

$$r_{iu} = g(s_i, a_i) + \eta_{iu}, \tag{1.4}$$

where each measurement error η_{iu} is a normal random variable with mean zero and variance $\sigma_\eta^2 >$

[4] See Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007).

[5] Note that the model allows for arbitrary differences between siblings in mean schooling levels and mean test scores. Such differences might arise, for instance, because of birth order effects on human capital investment as documented by Behrman and Taubman (1986) and Black et al. (2005).

0. The η_{iu} are assumed to be independent of each other and of all the other variables in the model.[6] The assumption that the errors in the productivity observations are uncorrelated between siblings plays no role in the analysis when learning is purely individual but simplifies the treatment of situations that involve social learning. Sections 1.2.3 and 1.2.4 discuss the relationship between wages and information sets under differing assumptions about the learning process.

The econometrician is assumed to observe a test score z_i in addition to the education level s_i. Letting $\theta_s > 0$ and $\theta_a > 0$, the ability measure z_i takes the form:

$$z_i = \theta_s s_i + \theta_a a_i + \omega_i, \tag{1.5}$$

where ω_i, which represents factors unrelated to labor productivity that affect the test score, is independent of both a_1, a_2 and s_1, s_2. The error terms ω_1, ω_2 are joint normally distributed with respective means $\mu_{\omega 1}$ and $\mu_{\omega 2}$, identical variance $\sigma_\omega^2 > 0$, and correlation $\rho_\omega \in (0, 1)$. In addition, the test score z_i is assumed to be unobservable to employers as in Altonji and Pierret (2001); so that, employers cannot use z_i as an additional signal of productivity when forming beliefs about a_i. Note that specification (1.5) resembles the model of the late test score in Griliches (1977). In particular, it accounts to some extent for the possibility that z_i is not a pure indicator of labor market ability a_i and is affected by human capital investment s_i.[7]

1.2.2 Informational Content of Test Scores and Schooling

In order to derive the empirical implications of wage determination under different learning processes, it is necessary to characterize the relationship among siblings' test scores, schooling choices, and innate abilities. Specifically, I analyze the coefficient obtained from a hypothetical regression of the siblings' abilities a_1, a_2 on their test scores z_1, z_2 after controlling for their schooling s_1, s_2. In the discussion that follows, I let σ_y^2 be the variance of the variable y_i and ρ_y be the correlation between y_1 and y_2. The analysis proceeds in two steps.

[6] Observe that the formulation in equation (1.4) is equivalent to one in which $g(s_i, a_i)$ is replaced with any linear function of a_i and s_i. None of the results derived here depend qualitatively on the assumption that the η_{iu} have identical variance or are serially uncorrelated.

[7] For example, Neal and Johnson (1996) present empirical evidence from the NLSY79 that additional schooling has a positive effect on AFQT scores.

The first step is to calculate the component of each sibling's test score that is orthogonal to her own and her sibling's schooling. The result below characterizes the problem of predicting the test scores z_1, z_2 from a regression on the schooling levels s_1, s_2.[8]

Proposition 1.2.1 *The regression coefficient of* $(z_1, z_2)'$ *on* $(s_1, s_2)'$ *is given by:*

$$\mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right]^{-1} = \theta_s \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta_a \frac{\gamma\sigma_a^2}{\sigma_s^2(1-\rho_s^2)} \begin{pmatrix} 1-\rho_a\rho_s & \rho_a-\rho_s \\ \rho_a-\rho_s & 1-\rho_a\rho_s \end{pmatrix}. \quad (1.6)$$

The first term in the above formula accounts for the causal effect of schooling on the test score.[9] The second term, which arises from the relationship of schooling with ability, is a generalization of the univariate measurement error formula, where schooling is treated as ability measured with error.[10]

Note that this formula provides a method for answering one of the central questions raised by Griliches (1979) regarding the role of families in human capital formation.[11] In particular, one can directly test whether the sibling correlation ρ_a in ability is greater than the sibling correlation ρ_s in schooling, especially if the schooling variable itself is not measured with error.[12] Formula (1.6) shows that if $(z_1, z_2)'$ is regressed on $(s_1, s_2)'$, then the coefficient on a sibling's schooling is positive if $\rho_a > \rho_s$ and negative if $\rho_a < \rho_s$.[13] As explained in Griliches (1979), this question is relevant when interpreting family fixed-effects estimates of the returns to schooling. Depending on whether siblings have a higher or lower correlation in ability than in schooling, the within-family estimator of the return to schooling may either mitigate or exacerbate ability bias relative to the ordinary least squares estimate of the schooling coefficient.

[8] The proof of the proposition is given in appendix A.1.

[9] See appendix A.3 for a further discussion of the endogeneity of the test score.

[10] Consider the special case where $\rho_a = \rho_s$ and $\gamma = 1$. Then, in the second term, the coefficient on one's own schooling is the parameter θ_a multiplied by the reliability ratio σ_a^2/σ_s^2. The coefficient on a sibling's schooling is zero in this case.

[11] An analogous formula applies if the log wage instead of the test score is used as the dependent variable, provided that the log wage can be modeled like the test score as a linear combination of ability, schooling, and an error term.

[12] The effect of classical measurement error in the schooling variable is examined in appendix A.9.

[13] Note that this procedure is more informative than simply comparing the raw sibling correlations in schooling and test scores, because it accounts for the fact that test scores are imperfect indicators of ability that contain measurement error whose correlation between siblings is unknown.

The intuition behind formula (1.6) is as follows. Suppose, for example, that $\rho_a < \rho_s$, in which case the model predicts a perverse negative relationship between a sibling's schooling and one's test score, conditional on one's own schooling. For a given realization of sibling 1's schooling s_1, a high value of sibling 2's schooling s_2 suggests that a_2 or ϵ_2 is high. Because $\rho_a < \rho_\epsilon$ whenever $\rho_a < \rho_s$, it follows that a_1 is less likely than ϵ_1 to be high when s_2 is high. Thus, holding constant the value of s_1, a high value of s_2 conveys unfavorable information about a_1. In other words, there are two determinants of the regression coefficient on a sibling's schooling after controlling for one's own schooling. A sibling's schooling reflects not only one's own ability but also other factors that affect one's schooling decision, such as discount rates, liquidity constraints, and institutional parameters. If schooling is more highly correlated among siblings than ability, then a sibling's schooling is more informative about these other influences on one's schooling than it is about one's ability.

Having calculated the component of each sibling's test score orthogonal to both siblings' schooling, the second step is to characterize the relationship between siblings' abilities and test scores after partialling out the influence of schooling. Denoting $s = (s_1, s_2)'$ and $z = (z_1, z_2)'$, consider the regression coefficient of the siblings' abilities on their schooling and test scores:

$$\mathbb{C}\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} s \\ z \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s \\ z \end{pmatrix}\right]^{-1} = \begin{pmatrix} \psi_o & \psi_f & \pi_o & \pi_f \\ \psi_f & \psi_o & \pi_f & \pi_o \end{pmatrix}. \tag{1.7}$$

The result below enumerates the basic properties of the regression parameters π_o and π_f, which represent the relationship of one's ability to one's own test score and a sibling's test score.[14]

Proposition 1.2.2 *The regression parameters π_o and π_f satisfy $\pi_o > \pi_f$, $\pi_o > 0$, and $\pi_o^2 > \pi_f^2$.*

The three parts of proposition 1.2.2 can be stated as follows. First, one's own test score remains a stronger predictor of one's ability than a sibling's test score after controlling for one's own and a sibling's schooling. Second, the partial correlation of one's ability with one's own test score is positive. Third, the coefficient on one's own test score is larger in absolute value than the coefficient on a sibling's test score in the regression of one's ability on one's own and a sibling's test scores and schooling. These simple properties will be important in deriving the empirical implications of

[14] The proof of the proposition is given in appendix A.1.

the learning models developed in the sections that follow.

Although proposition 1.2.2 demonstrates that one's own test score is positively related to one's ability given the other regressors, there is no analogous result for the coefficient on a sibling's test score, which can in general have either a positive or a negative partial correlation with one's ability. The reason for this ambiguity is that the test score is affected by factors other than ability that may be correlated between siblings. In other words, the sign of the coefficient π_f on a sibling's test score is the outcome of two competing effects: a positive correlation in ability a_i leads π_f to be positive, but a positive correlation in testing error ω_i leads π_f to be negative. This feature of the model is akin to the finding in proposition 1.2.1 that a sibling's schooling can have either a positive or a negative coefficient in the regression of one's test score on one's own and a sibling's schooling.[15]

The complexities introduced by such interactions in measurement error are the reason for assuming in section 1.2.1 that the error terms η_{1u_1} and η_{2u_2} in siblings' performance signals are uncorrelated. Although this assumption is seemingly inconsequential, it has substantive economic consequences for the analysis of social learning in section 1.2.4. If the measurement errors in performance signals were permitted to be correlated between siblings, then it would be possible for an individual's log wage to be negatively related to the performance of one's sibling. In other words, employers might assign a negative price to a sibling's performance when determining wages. This possibility arises because a high-performing sibling would indicate not only that one's own ability is high, but also that the measurement error in one's own performance signal is high. If the latter effect were to dominate, then a high-performing sibling would convey unfavorable information about one's own ability, conditional on one's own performance. The assumption of uncorrelated measurement error in the performance signals ensures that a sibling's performance enters an individual's log wage equation with a positive price.[16]

---

[15]Similarly, Altonji and Pierret (2001) find that the coefficient on the interaction term of each easy- or hard-to-observe variable with time cannot be signed in general when more than one variable of each type is included in their regression analysis, because each of these variables can have either a positive or a negative partial correlation with the component of employers' beliefs orthogonal to the information available when a worker first enters the labor market.

[16]Note, however, that this assumption plays no role in the analysis if employer learning has no social component. Hence, a failure to account for correlated measurement error in the performance signals cannot result in a rejection of the individual learning model if it is in fact true.

1.2.3 Individual Learning

This section analyzes the case where employer learning is individualistic; so that, the wage w_i of person $i \in \{1, 2\}$ is based only on her own education s_i and her own performance r_i. For concreteness, employers are assumed to set the wage w_i equal to the conditional expectation of labor productivity given s_i and r_i. The analysis here proceeds in two steps. I first express $\log(w_i)$ as a function of s_i and r_i. I then examine the conditional expectation of $\log(w_i)$ given s_1, s_2 and z_1, z_2.

In order to calculate $\log(w_i)$, I derive beliefs given s_i and r_i. Conditional on the observed level of schooling s_i, employers' beliefs about $g(s_i, a_i)$ are normally distributed with mean $\mu_{mi}(s_i)$ and variance σ_m^2 where:

$$\mu_{mi}(s_i) = \mathbb{E}[g(s_i, a_i)|s_i] = \beta s_i + \mathbb{E}(a_i|s_i) \;\text{ and }\; \sigma_m^2 = \mathbb{V}[g(s_i, a_i)|s_i] = \mathbb{V}(a_i|s_i). \tag{1.8}$$

Because all variables are assumed to be joint normally distributed, the conditional variance does not depend on the realization of any variable. From the results in DeGroot (1970), it follows that employers' beliefs about $g(s_i, a_i)$ given both s_i and r_i are normally distributed with mean $\mu_{gi}(s_i, r_i)$ and variance σ_{gi}^2 where:

$$\mu_{gi}(s_i, r_i) = (1 - \chi_i)\mu_{mi}(s_i) + \chi_i \bar{r}_i, \quad \sigma_{gi}^2 = (\sigma_m^{-2} + t_i \sigma_\eta^{-2})^{-1}, \quad \chi_i = t_i \sigma_\eta^{-2} \sigma_{gi}^2, \tag{1.9}$$

and $\bar{r}_i$ is the sample mean of the sequence $r_i = \{r_{iu}\}_{u=1}^{t_i}$. Note that equation (1.9) accounts for the effect of age on the precision of employers' beliefs about a worker's ability. As in Farber and Gibbons (1996) and Altonji and Pierret (2001), the precision $t_i \sigma_\eta^{-2}$ of the information contained in the sequence of performance observations increases with age t_i.[17] Thus, if worker 1 is older than worker 2, then χ_1 will be larger than χ_2, indicating that beliefs about worker 1 are based less on schooling and more on performance relative to beliefs about worker 2. The effect of differences in siblings' age levels on the relative precisions of employers' beliefs will play a central role in the model developed in section 1.2.4.

[17] One difference between the presentation here and in those papers is that individuals here accumulate productivity signals with age instead of with labor market experience. This expositional choice is justified if individuals acquire productivity signals not only when working but also while enrolled in school as suggested by Arcidiacono et al. (2010).

Given the normality of employers' beliefs about $g(s_i, a_i)$, the conditional expectation of labor productivity given s_i and r_i can be expressed as:

$$\mathbb{E}\{\exp[l(s_i, a_i, t_i)]|s_i, r_i\} = \exp[\mu_{gi}(s_i, r_i) + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i)]; \tag{1.10}$$

so that, the log wage of each person is simply:

$$\log(w_i) = \mu_{gi}(s_i, r_i) + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i), \tag{1.11}$$

where $\mu_{gi}(s_i, r_i)$ and σ_{gi}^2 are given by equation (1.9).

I now calculate the conditional expectation of the log wage given the information available to the econometrician. Using equations (1.4), (1.9), and (1.11), one obtains:

$$\mathbb{E}[\log(w_i)|s_1, s_2, z_1, z_2] = \chi_i \mathbb{E}(a_i|s_1, s_2, z_1, z_2) + f_i(s_i, t_i), \tag{1.12}$$

where the function $f_i(s_i, t_i)$ is given by:

$$f_i(s_i, t_i) = (1 - \chi_i)\mu_{mi}(s_i) + \chi_i \beta s_i + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i). \tag{1.13}$$

Note that the conditional expectation appearing on the right-hand side of equation (1.12) was the subject of proposition 1.2.2. In addition, equation (1.12) underscores the importance of controlling for an individual's own schooling when calculating the conditional expectation of the log wage. By doing so, the test scores z_1, z_2 appear in the conditional expectation only insofar as they predict the component of $\log(w_i)$ based on information about an individual's ability a_i gained from the productivity signal r_i. Because of this feature of the model, one can obtain the following invariance result concerning the ratio of coefficients on the test scores in a regression of $\log(w_1)$, $\log(w_2)$ on s_1, s_2 and z_1, z_2. The result is an immediate consequence of equation (1.12).

Proposition 1.2.3 *Suppose that learning is individual. Let α_{ij} denote the regression coefficient on person j's test score in the conditional expectation of person i's log wage given s_1, s_2 and z_1, z_2. Then $\alpha_{12}\alpha_{22} = \alpha_{21}\alpha_{11} = (\chi_1\pi_f)(\chi_2\pi_o)$.*

To understand this result, suppose that sibling 1 is older than sibling 2; so that, sibling 1's wage is based less on education and more on performance than sibling 2's wage. Because each sibling's wage has a different composition, it might be difficult to compare the results of wage regressions across siblings. The importance of proposition 1.2.3 is that it enables such comparisons to be made. Even though α_{11} is larger in magnitude than α_{22} by the proportion χ_1/χ_2, it follows from proposition 1.2.3 that α_{12} is also larger in magnitude than α_{21} by this proportion. In other words, the impact of a sibling's test score on one's log wage grows with age at the same rate as the impact of one's own test score. Because employers do not use information on a person's siblings in the current model, this result is also valid in the case where an individual has more than one sibling. Section 1.2.4 examines how deviations from this rule can arise when employers use information on one sibling when determining the wage of another sibling.

Nonetheless, it should be noted that there can exist situations in which employer learning is purely individual but the invariance result derived above does not hold. Suppose, for example, that unobserved ability is multidimensional instead of being the single factor assumed so far. Although this modification of the model is not problematic in itself, it now becomes possible to specify distinct learning processes for different components of ability; so that, employers learn more quickly about some components relative to others. Because these separate components are unlikely to all have the same relationship with schooling and test scores, there may no longer be a stable underlying association between test scores and log wages that holds across age levels. In such cases, it can be useful to compare siblings at the same age level instead of at the same point in time.[18]

1.2.4 Social Learning

This section examines the case in which employer learning has an element of statistical nepotism. In particular, the wage v_i of each sibling incorporates information on the education s_1, s_2 and performance r_1, r_2 of both siblings; so that, employers set the wage of sibling $i \in \{1, 2\}$ equal to the conditional expectation of her own labor productivity given s_1, s_2 and r_1, r_2.[19] Proceeding

[18] See appendix A.2 for an analysis of siblings at the same age level.

[19] A potential concern with this assumption about wage determination is that a sibling's characteristics may not be observable to a person's employer unless both individuals work for the same firm. To address this issue, appendix A.13 presents a simple model of employee referrals in which an older sibling's attributes can affect a younger sibling's log wage even if the two siblings work for different employers. The model there yields similar predictions to the setup here.

as in the previous section, I first derive the log wage as a function of the information available to employers and then compute the conditional expectation of the log wage given the variables observable to the econometrician.

I begin by calculating beliefs given s_1, s_2 and r_1, r_2. For a given sibling with index i, let e be the index of the other sibling; so that, $e = 2$ if $i = 1$, and vice versa. Conditional on the schooling s_i, s_e of both siblings and the performance r_e of sibling e, employers' beliefs about the time-invariant component $g(s_i, a_i)$ of sibling i's log productivity are normally distributed with mean $\mu_{ni}(s_i, s_e, r_e)$ and variance σ^2_{ni} where:

$$\mu_{ni}(s_i, s_e, r_e) = \beta s_i + \mathbb{E}(a_i|s_i, s_e, r_e) \text{ and } \sigma^2_{ni} = \mathbb{V}(a_i|s_i, s_e, r_e). \tag{1.14}$$

Note that the conditional variances σ^2_{ni}, σ^2_{ne} satisfy $\sigma^2_{ni} \geq \sigma^2_{ne}$ if $t_i \geq t_e$. This observation follows from the fact that the conditional expectation minimizes the mean squared error of the prediction; so that, as one controls for additional variables, the mean squared error cannot increase.[20] In other words, if sibling i is at least as old as sibling e, then sibling i's performance history is at least as long as sibling e's, which implies that r_i is no less informative about sibling e than r_e is about sibling i. Hence, beliefs about sibling e conditional only on s_e, s_i, and r_i are at least as precise as beliefs about sibling i conditional only on s_i, s_e, and r_e.

Because the measurement errors in the performance observations are independent of each other and of all the other variables in the model, employers' beliefs about $g(s_i, a_i)$ given both s_i, s_e and r_i, r_e are normally distributed with mean $\mu_{qi}(s_i, s_e, r_i, r_e)$ and variance σ^2_{qi} where:

$$\mu_{qi}(s_i, s_e, r_i, r_e) = (1 - \xi_i)\mu_{ni}(s_i, s_e, r_e) + \xi_i \bar{r}_i, \quad \sigma^2_{qi} = (\sigma^{-2}_{ni} + t_i \sigma^{-2}_{\eta})^{-1}, \quad \xi_i = t_i \sigma^{-2}_{\eta} \sigma^2_{qi}, \tag{1.15}$$

and $\bar{r}_i$ is the sample mean of $\{r_{iu}\}_{u=1}^{t_i}$. In equation (1.15), if t_i is greater than t_e, then $t_i \sigma^{-2}_{\eta}$ is greater than $t_e \sigma^{-2}_{\eta}$, and σ^{-2}_{ni} is no greater than σ^{-2}_{ne}; so that, ξ_i is greater than ξ_e. To paraphrase, if sibling i is older than sibling e, then beliefs about sibling i are based more on her own performance and less on other measures of her ability compared to beliefs about sibling e.

It follows that the conditional expectation of sibling i's labor productivity given s_i, s_e and r_i,

[20] Because all random variables are normally distributed, the mean squared error is identical to the conditional variance.

r_e is:

$$\mathbb{E}\{\exp[l(s_i, a_i, t_i)]|s_i, s_e, r_i, r_e\} = \exp[\mu_{qi}(s_i, s_e, r_i, r_e) + \tfrac{1}{2}\sigma_{qi}^2 + h(t_i)], \tag{1.16}$$

resulting in the proceeding expression for sibling i's log wage:

$$\log(v_i) = \mu_{qi}(s_i, s_e, r_i, r_e) + \tfrac{1}{2}\sigma_{qi}^2 + h(t_i), \tag{1.17}$$

where $\mu_{qi}(s_i, s_e, r_i, r_e)$ and σ_{qi}^2 are given by equation (1.15).

I now derive the conditional expectation of $\log(v_i)$ given s_i, s_e and z_i, z_e. Combining equations (1.4), (1.14), (1.15), and (1.17), one obtains:

$$\begin{aligned}\mathbb{E}[\log(v_i)|s_i, s_e, z_i, z_e] =& (1-\xi_i)\mathbb{E}[\mathbb{E}(a_i|s_i, s_e, r_e)|s_i, s_e, z_i, z_e] \\ &+ \xi_i\mathbb{E}(a_i|s_i, s_e, z_i, z_e) + b_i(s_i, t_i),\end{aligned} \tag{1.18}$$

where the function $b_i(s_i, t_i)$ is given by:

$$b_i(s_i, t_i) = \beta s_i + \tfrac{1}{2}\sigma_{qi}^2 + h(t_i). \tag{1.19}$$

Because the r_{ue} have identical covariances with each other and with s_i, s_e, and a_i, the conditional expectation of a_i given s_i, s_e, and r_e has the form:

$$\mathbb{E}(a_i|s_i, s_e, r_e) = \mathbb{E}(a_i|s_i, s_e, \bar{r}_e) = \zeta_{ci} + \zeta_{oi}s_i + \zeta_{fi}s_e + \zeta_{ri}\bar{r}_e; \tag{1.20}$$

so that, $(s_i, s_e, \bar{r}_e)$ is a sufficient statistic for (s_i, s_e, r_e) with respect to a_i. Thus, the iterated expectation in equation (1.18) can be expressed as:

$$\mathbb{E}[\mathbb{E}(a_i|s_i, s_e, r_e)|s_i, s_e, z_i, z_e] = \zeta_{ri}\mathbb{E}(a_e|s_i, s_e, z_i, z_e) + d_i(s_i, s_e), \tag{1.21}$$

where $d_i(s_i, s_e)$ is defined as:

$$d_i(s_i, s_e) = \zeta_{ci} + \zeta_{oi}s_i + (\zeta_{fi} + \zeta_{ri}\beta)s_e. \tag{1.22}$$

Combining equations (1.18) and (1.21), I obtain the final expression for the conditional expectation of the log wage:

$$\mathbb{E}[\log(v_i)|s_i, s_e, z_i, z_e] = (1-\xi_i)\zeta_{ri}\mathbb{E}(a_e|s_i, s_e, z_i, z_e) + \xi_i\mathbb{E}(a_i|s_i, s_e, z_i, z_e) + p_i(s_i, s_e, t_i), \quad (1.23)$$

where $p_i(s_i, s_e, t_i)$ is given by:

$$p_i(s_i, s_e, t_i) = b_i(s_i, t_i) + (1 - \xi_i)d_i(s_i, s_e). \quad (1.24)$$

Equation (1.23) demonstrates that in the presence of social interactions in the learning process, the log wage can be decomposed into two separate components, one of which contains a person's own ability, and the other of which reflects her sibling's ability. Note that in the previous section where employer learning was individual, it was essential to control for one's own schooling, in order to identify the relationship of test scores to the component of the wage based on one's own ability. In the current model, the wage also incorporates information on one's sibling, making it important to control as well for the sibling's schooling, so as to isolate the part of the log wage based on the abilities of the two siblings.

It is now possible to prove the following counterpart to proposition 1.2.3 for the current model in which employer learning has a social component.[21] The first part of the proposition is an immediate consequence of the symmetric treatment of the two siblings.[22] In the second part of the proposition, an analogous statement holds if $t_2 > t_1$ instead of $t_1 > t_2$.

Proposition 1.2.4 *Suppose that learning is social. Let ν_{ij} denote the regression coefficient on person j's test score in the conditional expectation of person i's log wage given s_1, s_2 and z_1, z_2.*

1. *If $t_1 = t_2$, then $\nu_{12}\nu_{22} = \nu_{21}\nu_{11}$.*

2. *If $t_1 > t_2$, then $\nu_{12}\nu_{22} < \nu_{21}\nu_{11}$.*

On the one hand, the first part of proposition 1.2.4 can be regarded as a variant of the results in Manski (1993) concerning the difficulties of distinguishing between social and nonsocial effects. If

[21] The proof of the proposition is given in appendix A.1.

[22] To be specific, if $t_1 = t_2$, then $\xi_1 = \xi_2$ in equation (1.15); so that, the conditional expectations of $\log(v_1)$, $\log(v_2)$ in equation (1.18) have the same form up to a constant.

the two siblings have the same age, then both the individual and the social learning model predict that the ratio of the coefficients on test scores should be the same for the two siblings. On the other hand, if there are asymmetries in the age levels of the siblings, then the two models generate different predictions regarding the relative values of this ratio, making it possible to detect social interactions in the learning process.

Intuitively, if sibling 1 is older than sibling 2, then the labor market acquires more precise information about sibling 1 than about sibling 2, because sibling 1's performance is observed over a longer length of time than sibling 2's. Thus, when the market forms Bayesian beliefs about the abilities of the two siblings, greater weight is placed on information about sibling 1 than on information about sibling 2. Given that the labor market is competitive, this relative weighting is reflected in the wages of the two siblings; so that, the component of sibling 2's wage attributable to sibling 1's ability is larger than the component of sibling 1's wage attributable to sibling 2's ability. This phenomenon manifests itself in the data available to the econometrician as follows. If sibling 1 is older than sibling 2, then the ratio of the coefficient on a sibling's test score to the coefficient on one's own test score is typically higher in sibling 2's log wage than in sibling 1's log wage, because sibling 2's log wage is more strongly influenced by sibling 1's ability than vice versa, and sibling 1's ability is more strongly associated with sibling 1's than with sibling 2's test score, conditional on the other regressors.

1.3 Extensions of Employer Learning Models

The supplemental appendices discuss several extensions of the employer learning models developed in the preceding section. These extensions demonstrate the robustness of the theoretical predictions to simple changes in the setup of the model. Appendix A.2 describes the relationship between log wages and test scores when younger and older siblings are compared at a given age level instead of in a given time period. Appendix A.3 examines the issues that arise when the test score is endogenously influenced by an individual's schooling at the time of taking the test. Appendix A.4 explains how the analysis of social learning can be generalized to allow for an arbitrary number of siblings in a family. Appendix A.5 studies a social learning model in which an individual has at least two siblings with one sibling being older than the other.

1.4 Empirical Implementation

This section addresses some issues concerning the estimation of the learning models in section 1.2. A potential obstacle to implementing the tests in propositions 1.2.3 and 1.2.4 is that the regression coefficients are predicted to change with age in the conditional expectation of one's log wage given one's own and a sibling's test scores and schooling.[23] The analysis in section 1.2 treated the ages of the siblings in each family as being fixed. In the data, however, siblings from a sample of households are interviewed over multiple years, and the age structure varies across families and over time; so that, the coefficients in the conditional expectation function may not be constant across different households and survey years. One way to deal with this problem might simply be to include interactions of schooling and test scores with age when estimating the conditional expectation function. Nonetheless, this approach is unattractive in the current setting, because the social learning model implies that the coefficients on test scores are a function not only of one own's age but also of a sibling's age. Hence, the number of interaction terms that would need to be included in the specification is an order of magnitude greater than that required under individual learning, making it difficult to obtain precise estimates for the coefficients of interest given the limited data at disposal.

Remarkably, there is a simple procedure that in large part overcomes this estimation problem. First, I show that the main predictions of both the individual and the social learning model hold in aggregate. Specifically, if one considers all the pairs of younger and older siblings that can be derived from a sample of sibships with different age structures, then the predictions of each of the two learning models regarding the coefficients on test scores in a log wage regression also apply to the expected values of these coefficients for a randomly selected pair of siblings. This finding is somewhat surprising because these predictions involve a nonlinear function of the regression coefficients: the ratio of the coefficient on a sibling's test score to that on one's own test score. Nevertheless, the normality assumptions in this paper impose sufficient structure on the learning process to make aggregation of this sort possible. Second, I show that the pooled ordinary least squares estimator of the conditional expectation function will under reasonable conditions generate a consistent estimate of the expected values of the regression coefficients for a randomly selected

[23]In particular, the regression coefficients will depend on only one's own age under individual learning and on both one's own and a sibling's ages under social learning.

pair of siblings, provided that one controls sufficiently flexibly for the ages of the siblings under analysis. To simplify the exposition in this section, I assume that all families consist of exactly two siblings and that all sibships enter the labor market in the same year.[24]

Consider a random sample of $I \geq 1$ sibships. The families in the sample are indexed from 1 to I, and the siblings in each family are labeled 1 and 2. Sibling 1 is assumed to be older than sibling 2. There are D years under observation, which are labeled from 1 to D. Both members of each sibship i are assumed to be working in all of these years. Let $t_{i,j,d}$ represent the age of sibling j from family i in year d, and let $s_{i,j}$ and $z_{i,j}$ respectively denote the schooling and the test score of sibling j from family i. The age of each person increases by one in each year. Letting $t_{i,0} = (t_{i,1,0}, t_{i,2,0})$ represent the ages of the two siblings from family i in year zero, the set T of possible realizations of $t_{i,0}$ is taken to be finite. Every element of T is assumed to be a pair of distinct nonnegative integers.

Let $b_{i,j}$ be a $K \times 1$ vector of background variables for sibling j from family i. Although these variables were not discussed earlier, there is a simple way to formally introduce them into the framework in section 1.2.1 without changing the predictions of either learning model.[25] Assuming that $b_{i,j}$ is observable both to employers and to the econometrician, let the respective means $\mu_{a,i,j}$, $\mu_{\epsilon,i,j}$, $\mu_{\omega,i,j}$ of $a_{i,j}$, $\epsilon_{i,j}$, $\omega_{i,j}$ have the form:

$$\begin{aligned}(\mu_{a,i,j}, \mu_{\epsilon,i,j}, \mu_{\omega,i,j}) &= \mathbb{E}\left[(a_{i,j}, \epsilon_{i,j}, \omega_{i,j})|b_{i,1}, b_{i,2}, t_{i,0}\right] \\ &= \left(\phi_{a,0} + b'_{i,j}\phi_a,\ \phi_{\epsilon,0} + b'_{i,j}\phi_\epsilon,\ \phi_{\omega,0} + b'_{i,j}\phi_\omega\right),\end{aligned} \tag{1.25}$$

where $\phi_{a,0}$, $\phi_{\epsilon,0}$, and $\phi_{\omega,0}$ are constants, and ϕ_a, ϕ_ϵ, and ϕ_ω are $K \times 1$ coefficient vectors.[26]

Each sibling pair can be represented by the triple (i, p, q), where i indexes the family from which the two siblings are drawn, and p and q are the respective labels of the first and the second

[24] Appendix A.8 contains a more general treatment that extends the setup here to allow for an arbitrary number of siblings in a family and for different dates of labor market entry across families.

[25] Pinkston (2009) uses a similar strategy to add demographic characteristics to a model of asymmetric employer learning.

[26] The other parameters of the model in section 1.2.1—β, σ_a^2, ρ_a, γ, σ_ϵ^2, θ_a, θ_s, σ_ω^2, ρ_ω, σ_η^2—are assumed not to depend on the realizations of $b_{i,1}$, $b_{i,2}$, and $t_{i,0}$. The term $h(t_{i,j,d})$ in equation (1.1) is assumed to be a function only of $t_{i,j,d}$.

siblings in the pair.[27] I define two vectors:

$$t_{i,(p,q),d} = (t_{i,p,d}, t_{i,q,d})', \quad x_{i,(p,q)} = (z_{i,p}, z_{i,q}, s_{i,p}, s_{i,q}, b'_{i,p}, b'_{i,q})', \tag{1.26}$$

where $t_{i,(p,q),d}$ represents the ages of the two siblings from family i in year d, and $x_{i,(p,q)}$ contains their labor market characteristics. The conditional expectation of the log wage $w_{i,p,d}$ of sibling p from family i in year d given $x_{i,(p,q)}$ and $t_{i,(p,q),d}$ can be put in the following general form both under individual and under social learning:

$$\mathbb{E}\left(w_{i,p,d} | x_{i,(p,q)}, t_{i,(p,q),d}\right) = c(t_{i,(p,q),d}) + x'_{i,(p,q)} v(t_{i,(p,q),d}), \tag{1.27}$$

where $v(t_{i,(p,q),d})$ is a $(2K+4) \times 1$ coefficient vector, and $c(t_{i,(p,q),d})$ is a constant.[28] Note that $v(t_{i,(p,q),d})$ and $c(t_{i,(p,q),d})$ can vary with the age vector $t_{i,(p,q),d}$ of the two siblings from family i in year d.

I next define the two parameters of interest. For each family i, let G_i be a random variable that takes on the value of each natural number between 1 and D with equal probability $1/D$. Each realization of G_i represents a particular year from the set of observed dates. The random variable G_i is assumed to be independent of all the other variables in the model. Letting $\delta(\tilde{t}_{i,0})$ denote the proportion of families in which the ages of the two siblings in year zero are $\tilde{t}_{i,0} \in T$, the expected value ν_H of $v(t_{i,(1,2),G_i})$ is equal to:

$$\nu_H = \mathbb{E}[v(t_{i,(1,2),G_i})] = D^{-1} \sum_{\tilde{t}_{i,0} \in T} \delta(\tilde{t}_{i,0}) \sum_{d=1}^{D} v(\tilde{t}_{i,(1,2),0} + d \cdot \mathbb{1}_2), \tag{1.28}$$

and the expected value ν_L of $v(t_{i,(2,1),G_i})$ is equal to:

$$\nu_L = \mathbb{E}[v(t_{i,(2,1),G_i})] = D^{-1} \sum_{\tilde{t}_{i,0} \in T} \delta(\tilde{t}_{i,0}) \sum_{d=1}^{D} v(\tilde{t}_{i,(2,1),0} + d \cdot \mathbb{1}_2), \tag{1.29}$$

---

[27] Note that each family i contains two sibling pairs: $(i, 1, 2)$ and $(i, 2, 1)$.

[28] The expression for the conditional expectation in equation (1.27) is a consequence of equation (1.12) if learning is individual and of equation (1.23) if learning is social.

where $\mathbb{1}_2$ is a 2×1 vector of ones. For a randomly sampled family, ν_H and ν_L can be interpreted as the average values of the coefficient vectors $v(t_{i,(1,2),G_i})$ and $v(t_{i,(2,1),G_i})$ in a random year.[29]

It is now possible to state the following result, which is a generalization of propositions 1.2.3 and 1.2.4.[30] Consider the conditional expectation function in equation (1.27) as well as the expected values of the coefficient vectors in equations (1.28) and (1.29). First, if learning is individual, then the ratio of the second to the first entry of ν_H will be equal to the ratio of the second to the first entry of ν_L. That is, under individual learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will be the same as the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage. Second, if learning is social, then the ratio of the second to the first entry of ν_H will be less than the ratio of the second to the first entry of ν_L, especially assuming that the first entries of ν_H and ν_L are both positive. That is, under social learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will typically be lower than the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage.

Proposition 1.4.1 *For $i \in \{1, 2\}$, let $\nu_{H,i}$ denote the i^{th} element of the vector ν_H in equation (1.28), and let $\nu_{L,i}$ denote the i^{th} element of the vector ν_L in equation (1.29).*

1. *If learning is individual, then $\nu_{H,2}\nu_{L,1} = \nu_{L,2}\nu_{H,1}$.*

2. *If learning is social, then $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$.*

Having shown that the main predictions of the learning models survive aggregation, I now discuss the estimation of the expected values ν_H and ν_L of the coefficient vectors $v(t_{i,(1,2),G_i})$ and $v(t_{i,(2,1),G_i})$. Fixing any nonnegative integer M, let P represent the set composed of every pair of nonnegative integers whose sum is no greater than M. Letting $\#P$ be the number of elements

[29] Observe that the first and second elements of the vector ν_H (resp. ν_L) represent the average values of the coefficients on one's own and a younger (resp. an older) sibling's test scores in the conditional expectation of an older (resp. a younger) sibling's log wage in equation (1.27).

[30] The proof of proposition 1.4.1 is omitted from the paper, because a more general version of the result is proved in appendix A.8.

in the set P, the elements of P can be labeled from 1 to $\#P$ with $e^s = (e_1^s, e_2^s)$ denoting the s^{th} element of P. Given a 2×1 vector $t = (t_1, t_2)'$, let f_t denote the $\#P \times 1$ vector whose s^{th} entry is equal to the product $t_1^{e_1^s} t_2^{e_2^s}$; so that, f_t consists of one element for every term of a M^{th}-order bivariate polynomial in t. Let $h_{i,(p,q),d}$ be the $(2K + 4 + \#P) \times 1$ vector formed by stacking the vector $x_{i,(p,q)}$ on top of the vector $f_{t_{i,(p,q),d}}$. That is, I define:

$$h_{i,(p,q),d} = (x'_{i,(p,q)}, f'_{t_{i,(p,q),d}})', \tag{1.30}$$

where $x_{i,(p,q)}$ comprises the test scores, schooling levels, and background attributes of the two siblings from family i, and $f_{t_{i,(p,q),d}}$ contains the terms of a bivariate polynomial in their ages in year d.

Some further assumptions become relevant when estimating ν_H and ν_L. Fix $(p, q) = (1, 2)$ or $(p, q) = (2, 1)$. First, the conditional expectation of $x_{i,(p,q)}$ given that $t_{i,(p,q),G_i} = t$ is assumed to be adequately approximated by a M^{th}-order bivariate polynomial in t. That is, I assume that:

$$\mathbb{E}(x_{i,(p,q)} | t_{i,(p,q),G_i} = t) = \sum_{e \in P} \alpha^e_{(p,q)} (t_1^{e_1} t_2^{e_2}), \tag{1.31}$$

where t is any 2×1 vector of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability, and $\alpha^e_{(p,q)}$ is a $(2K + 4) \times 1$ vector that does not depend on t. Second, the matrix representing the expected value of $h_{i,(p,q),G_i} h'_{i,(p,q),G_i}$ is required to be nonsingular. That is, I assume that:

$$\text{rank}\big[\mathbb{E}(h_{i,(p,q),G_i} h'_{i,(p,q),G_i})\big] = 2K + 4 + \#P. \tag{1.32}$$

Third, the variance of $x_{i,(p,q)}$ given that $t_{i,(p,q),G_i} = t$ is restricted to be a matrix of constants that do not vary with t. That is, letting $r_{i,(p,q),G_i} = x_{i,(p,q)} - \mathbb{E}(x_{i,(p,q)} | t_{i,(p,q),G_i})$, I assume that:

$$\mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i} | t_{i,(p,q),G_i} = t) = \Sigma_{x,(p,q)}, \tag{1.33}$$

where t is any 2×1 vector of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability, and $\Sigma_{x,(p,q)}$ is a $(2K + 4) \times (2K + 4)$ matrix of constants that do not depend on t.[31] In addition,

[31] This restriction on the conditional variance matrix can be weakened to some extent. Specifically,

note that all random variables are treated as having finite first and second moments.

The following result shows that, under the assumptions above, the parameters ν_H and ν_L can be consistently estimated simply by pooling the observations on each sibling pair across every year and running ordinary least squares regressions on the resulting dataset. In particular, let:

$$\tilde{\nu}_H = \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(1,2),d} h'_{i,(1,2),d} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(1,2),d} w_{i,1,d} \right), \tag{1.34}$$

and let:

$$\tilde{\nu}_L = \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(2,1),d} h'_{i,(2,1),d} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(2,1),d} w_{i,2,d} \right). \tag{1.35}$$

Let $\hat{\nu}_H$ and $\hat{\nu}_L$ be vectors containing the first $2K+4$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$, respectively. That is, $\hat{\nu}_H$ (resp. $\hat{\nu}_L$) denotes the estimated coefficient on the covariate vector $x_{i,(1,2)}$ (resp. $x_{i,(2,1)}$) in a log wage regression that also controls for $f_{t_{i,(1,2),d}}$ (resp. $f_{t_{i,(2,1),d}}$). The result below shows that as the number of sampled sibships I goes to infinity, the estimators $\hat{\nu}_H$ and $\hat{\nu}_L$ converge in probability to ν_H and ν_L, respectively. The proof of proposition 1.4.2 is similar to those given by Wooldridge (2004), who examines the identification of average partial effects in correlated random coefficient models.[32]

Proposition 1.4.2 *Suppose that the assumptions in equations (1.31), (1.32), and (1.33) are satisfied. As the number of sampled sibships I goes to infinity, the estimators $\hat{\nu}_H$ and $\hat{\nu}_L$, which consist of the first $2K+4$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$ in equations (1.34) and (1.35), respectively converge in probability to ν_H and ν_L, which are defined in equations (1.28) and (1.29).*

The results in propositions 1.4.1 and 1.4.2 suggest the following basic strategy for estimating and testing the individual and social learning models.[33] The log wages of all older (resp. younger)

---

proposition 1.4.2 remains valid if equation (1.33) is replaced by $\mathbb{E}[\Sigma_{x,(p,q)}(t_{i,(p,q),G_i})v(t_{i,(p,q),G_i})] = \mathbb{E}[\Sigma_{x,(p,q)}(t_{i,(p,q),G_i})]\mathbb{E}[v(t_{i,(p,q),G_i})]$, where $\Sigma_{x,(p,q)}(t) = \mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i} | t_{i,(p,q),G_i} = t)$ for any 2×1 vector t of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability. That is, the random coefficient vector $v(t_{i,(p,q),G_i})$ is assumed to be uncorrelated with the random conditional variance matrix $\Sigma_{x,(p,q)}(t_{i,(p,q),G_i})$.

[32] The proof of proposition 1.4.2 is omitted from the paper, because a more general version of the result is proved in appendix A.8.

[33] See sections 1.5 and 1.6 for more detailed information regarding the construction of the estimation sample and the specification of the learning models.

siblings across all years are regressed both on their own and their younger (resp. older) sibling's test scores, schooling levels, and background attributes and on a bivariate polynomial in their own and their younger (resp. older) sibling's ages. When computing standard errors and test statistics, the Huber-White estimator of the variance-covariance matrix is used to allow for arbitrary forms of correlation among the error terms of observations on the same family. To evaluate the nonlinear restriction implied by the individual learning model in proposition 1.4.1, I calculate the Wald statistic for the null hypothesis that the coefficient on a younger sibling's test score in an older sibling's log wage times the coefficient on one's own test score in a younger sibling's log wage minus the coefficient on an older sibling's test score in a younger sibling's log wage times the coefficient on one's own test score in an older sibling's log wage is equal to zero.[34]

1.5 Dataset Construction and Description

The dataset used in this study is constructed from the 1979-2008 waves of the National Longitudinal Survey of Youth 1979 (NLSY79), which contains panel data on 12,686 men and women aged 14-22 in 1979. Respondents were interviewed annually from 1979 to 1994 and biennially thereafter. The NLSY79 is especially well suited to the purpose of this paper, which is to examine social learning among siblings in the labor market. Because the Armed Services Vocational Aptitude Battery (ASVAB) was administered to participants in the NLSY79, a growing literature on employer learning uses the resulting Armed Forces Qualification Test (AFQT) score as an ability measure that is not directly observable to employers. In addition, a large number of sibling studies analyze data from the NLSY79, which includes 5,863 respondents who have one or more interviewed siblings. The current paper uses the AFQT scores of pairs of siblings to identify the components of a person's log wage based on information about one's own and a sibling's ability.

In order to implement the empirical strategy developed above, I assemble a dataset in which each observation represents a particular sibling pair in a given survey year.[35] This dataset will serve

[34] By the delta method, this test statistic is, in general, asymptotically distributed as chi-squared with one degree of freedom.

[35] If there are two siblings p and q, then a sibling pair in which sibling p appears first and sibling q second is regarded as distinct from a sibling pair in which sibling q appears first and sibling p second. For example, if a family consists of three siblings, then six different pairs of siblings can be formed.

as the main estimation sample for the paper; therefore, the current section describes in detail how this dataset is constructed.[36] However, a variety of other related samples are used for the empirical analysis in certain sections, and the selection criteria for these samples are provided in the relevant text and tables. The data are derived from the 6,111 respondents in the cross-sectional sample and the 5,295 respondents in the supplemental sample of the NLSY79.[37] For this group of 11,406 individuals, I identify every survey year in which a respondent has a non-missing wage observation on a full-time job, where full-time is defined as 35 or more hours per week.[38] Each wage is then deflated using the CPI to a base period of 1982-1984, and any real wage less than $1 or greater than $100 is omitted from the analysis. The resulting sample consists of 126,101 observations on 10,945 individuals.

I next compile the required information on each respondent's education and AFQT score. Schooling is defined in each survey year as the highest grade that a respondent reports having completed up to the time of the current interview.[39] The AFQT scores are standardized among all respondents in the NLSY79 with the same year of birth, so as to account directly for the effect of age on one's test score. An additional 5,339 observations from 603 individuals with missing data on education or AFQT scores are deleted at this stage. I also exclude all observations that occur prior to the first survey year in which a respondent has left school for the first time. This criterion reduces the sample size to 117,649 observations on 10,331 individuals. In restricting the estimation sample to respondents who have left school for the first time and who are working at a full-time job, I seek to identify those individuals who are well attached to the workforce and for whom employer learning is likely to be relevant.[40]

[36]The main estimation sample is restricted to observations on sibling pairs in which both members have worked since the last interview. However, selection into employment may not be entirely exogenous. Therefore, appendix A.21 expands the dataset to include non-working individuals and examines the joint work-wage outcomes of respondents. I continue to find evidence of sibling effects on labor market variables after performing this extension.

[37]The military sample does not contain any pairs of interviewed siblings; therefore, all 1,280 of its members are excluded from the empirical analysis.

[38]If an individual reports holding multiple jobs, then only information on the CPS job is used.

[39]As in Altonji and Pierret (2001), schooling is required to be nondecreasing over time. Hence, if an individual reports a higher schooling level in one year and a lower schooling level in a later year, then the education variable is set equal to the greater of these two values in both years.

[40]Nevertheless, the empirical results in Tables 1.3 and 1.5 do not change substantially if the estimation sample is expanded to include respondents who have not yet left school or who work at a part-time job.

Having created a sample of wage observations on the period after each individual's initial transition from school to work, I isolate those respondents for whom valid sibling data is available. There are 7,639 observations from 623 individuals who are missing data on either their total number of siblings or their number of older siblings.[41] These observations are excluded from the estimation sample, bringing the sample size to 110,010 observations on 9,708 individuals. At this point, each observation in the sample corresponds to a specific person in a given survey year. To generate a new dataset consisting of sibling pairs instead of individuals, I apply the following procedure to the existing sample. For each person in the sample in a given survey year, I search over all the other individuals in the sample in that year for an observation on the person's sibling. If a sibling is found, then an observation containing information on the person and her sibling is added to the new dataset in that survey year. In the event that multiple siblings are located, one such observation is included for each sibling identified. If no sibling exists, then the person is excluded from the main estimation sample in that year. The resulting dataset contains 55,242 observations on 7,180 sibling pairs, which comprise 4,766 individuals from 2,012 sibships. Because the empirical strategy is based primarily on differences in age between siblings, any pairs of siblings in which both members have the same year and month of birth are eliminated, which reduces the dataset to 54,474 observations on 7,074 sibling pairs, covering 4,726 individuals from 1,993 families.[42]

Table 1.1 presents descriptive statistics for the main sample used in the empirical analysis. 58.88 percent of observations in the data originate from households in the nationally representative cross-sectional sample of the NLSY79, with the remaining 41.12 percent of observations belonging to members of the supplemental sample, which oversamples minorities and disadvantaged whites.[43] The demographic composition of the estimation sample is 42.43 percent female, 32.22 percent black, and 17.71 percent Hispanic. In locational terms, 77.70 percent of observations are on respondents living in an urban area, and 41.58 percent of observations are on individuals residing in the South. The sample is composed primarily of siblings from large families with the

[41] These figures include 3,823 observations from 300 individuals who do not have data on their number of older siblings because they are only children.

[42] The analysis in Table 1.11 does not exploit age differences between siblings; therefore, twins are not excluded from the results reported there.

[43] Because the background characteristics of the sample used in this paper differ from those of the workforce at large, the empirical results in Tables 1.3 and 1.5 were replicated using only those sibling pairs belonging to the representative cross-sectional sample. The findings were similar to those reported in the paper.

average sibship size being 5.732 across all observations.[44] The mean birth order is 2.849 for the older member and 4.203 for the younger member of each sibling pair. The corresponding difference in age between the older and the younger sibling in each pair is on average 2.53 years.[45]

Turning to the labor market outcomes of siblings in the estimation sample, the average hourly real wage is \$8.10 for the older sibling and \$7.57 for the younger sibling in each pair. Sample members appear to have a substantial commitment to the labor force, working an average of 43.10 hours per week across all observations. The mean schooling level is 12.89, and the mean AFQT score is -0.0566.[46] The average differences between older and younger siblings in these human capital measures are not large. The older sibling in each pair typically has 0.12 more years of schooling than the younger sibling. However, after standardizing the AFQT scores within each birth cohort, younger siblings have AFQT scores that are on average 0.0685 standard deviations higher than those of older siblings. Therefore, the 6.23 percent higher wage earned by older relative to younger siblings appears to be attributable mostly to greater experience as opposed to underlying disparities in educational attainment or cognitive ability.[47]

For the purpose of identifying social interactions in employer learning, it is desirable for there to be a moderate correlation in ability among siblings. On the one hand, if the sibling correlation in ability is close to zero, then a sibling's performance is a weak predictor of one's own ability and has little impact on one's log wage. On the other hand, if the sibling correlation in ability is near perfect, then each person's test score has a similar relationship to both one's own and a sibling's ability; so that, the components of a person's log wage based on one's own and a

[44]Note that larger sibships tend to be overrepresented in the estimation sample because each observation from a given survey year represents a sibling pair instead of an individual. For example, a sibship with two members contains only two pairs of siblings, whereas a sibship with three members consists of six sibling pairs.

[45]Because the employer learning process is relatively rapid as demonstrated by Lange (2007), even an age gap of 2.53 years can generate a sizeable disparity in the precision of the market's beliefs about each sibling's ability, especially at the early stages of a worker's career. The estimates in Lange (2007) indicate that the variance of the market's ability prediction error declines by 41.17 percent after two years and 51.21 percent after three years of employment.

[46]Since AFQT scores are standardized using all respondents in the NLSY79 instead of only members of the estimation sample, the mean and standard deviation of the AFQT scores reported in Table 1.1 differ slightly from zero and one, respectively.

[47]Nevertheless, to the extent that such birth order effects are present in the data, the framework in section 1.2 accounts for any differences between younger and older siblings in the mean levels of schooling, test scores, and productive ability.

Table 1.1: Summary Statistics for Sibling Pairs in Labor Market

	Both Siblings	Older Sibling	Younger Sibling
Labor Market			
Real Hourly Wage			
Mean (S.D.)	7.838 (5.336)	8.101 (5.565)	7.574 (5.082)
Inter-Sib. Corr.	0.3556	——	——
Log Real Hourly Wage			
Mean (S.D.)	1.908 (0.527)	1.939 (0.531)	1.877 (0.522)
Inter-Sib. Corr.	0.3882	——	——
Weekly Hours Worked			
Mean (S.D.)	43.10 (7.46)	43.23 (7.55)	42.96 (7.37)
Human Capital			
Years of Schooling			
Mean (S.D.)	12.89 (2.40)	12.95 (2.41)	12.83 (2.38)
Inter-Sib. Corr.	0.5174	——	——
Standardized AFQT			
Mean (S.D.)	-0.0566 (1.0019)	-0.0909 (1.0150)	-0.0224 (0.9875)
Inter-Sib. Corr.	0.6586	——	——
Basic Demographics			
Pct. Black	32.22	——	——
Pct. Hispanic	17.71	——	——
Pct. Female	42.43	42.30	42.55
Pct. Urban	77.70	78.00	77.40
Region of Residence			
Pct. Northcentral	25.09	25.14	25.04
Pct. Northeast	16.49	16.64	16.34
Pct. South	41.58	41.41	41.75
Pct. West	16.85	16.82	16.88
Age			
Mean (S.D.)	31.07 (7.34)	32.34 (7.25)	29.81 (7.21)
Family Background			
Birth Order			
Mean (S.D.)	3.526 (2.231)	2.849 (2.075)	4.203 (2.175)
Sibship Size			
Mean (S.D.)	5.732 (2.657)	——	——
Mother's Schooling			
Mean (S.D.)	10.78 (3.14)	——	——
Father's Schooling			
Mean (S.D.)	10.67 (4.08)	——	——
Mother's Age			
Mean (S.D.)	57.38 (9.02)	——	——
Father's Age			
Mean (S.D.)	60.95 (9.66)	——	——
Sample Size			
No. Families	1993	——	——
No. Individuals	4726	2695	2707
No. Sibling Pairs	7074	——	——
No. Observations	54474	——	——

Note: The hourly wage is deflated using the CPI with 1982-1984 as the base period. The reference group for standardizing the AFQT score is all respondents in the NLSY79 having the same year of birth. Observations with missing data on a given variable are omitted when calculating the summary statistics for that variable.

sibling's performance cannot be empirically distinguished from each other.[48] Using the available human capital measures as indicators of ability, the sibling correlations in AFQT scores, schooling, and log wages are 0.6586, 0.5174, and 0.3882, respectively.[49] For comparison, the survey of sibling studies in Griliches (1979) lists the following ranges of intraclass correlation coefficients between non-twin brothers: 0.47-0.58 for test scores, 0.24-0.55 for schooling, and 0.11-0.40 for log earnings.

1.6 Empirical Results

This section presents empirical evidence of social interactions among siblings in the labor market. Section 1.6.1 documents the influence of sibship size and birth order on job search behavior. Section 1.6.2 analyzes the relationships among the AFQT scores, schooling levels, and log wages of younger and older siblings. Section 1.6.3 performs various robustness checks on the main estimation results, such as analyzing the impacts of youngest and oldest siblings on middle siblings, modifying the empirical specification to include interactions among more than two siblings, comparing the log wages of older and younger siblings at the same age, and accounting for the causal effect of schooling on AFQT scores. Section 1.6.4 tests additional predictions of the learning models for the coefficients on one's own and a sibling's schooling in log wage equations and for the variance of the change in log wage residuals between successive age levels. Section 1.6.5 conducts three falsification exercises based on samples of sibling pairs in which one member has little labor market experience, was geographically isolated from the other member, or moves to a different region.

[48]In order to assess whether the variation in ability among siblings is sufficient to differentiate between the components of one's log wage based on one's own and a sibling's ability, family fixed-effects estimates of the impact of the AFQT score on the log wage are performed in appendix A.14.

[49]The sibling correlation in log wages is based on the real hourly wage in each survey year. As discussed in Solon et al. (1991), the sibling correlation in permanent earnings is likely to be greater than the correlation between siblings in a single-year measure of earnings. Averaging the log wage of each member of a sibling pair over all the survey years in which that sibling pair appears in the main estimation sample, the sibling correlation in log wages rises to 0.4157.

1.6.1 Job Search Estimates

Job search is a potential channel through which employers might acquire information about a person's siblings. This section, therefore, analyzes the job search patterns of siblings as background evidence for the main empirical results on social learning.[50] The first four columns in the upper panel of Table 1.2 report estimates for linear probability models relating birth order and sibship size to the likelihood of obtaining one's most recent job with the help of a sibling. In a parsimonious specification that controls only for a gender dummy and fixed effects for year of birth, the coefficients on sibship size and birth order are respectively 0.0087 and 0.0133 with corresponding standard errors 0.0014 and 0.0017 when each variable is used individually as a regressor. However, when both variables are jointly included in the regression, the coefficient of 0.0019 on sibship size is not significantly different from zero, whereas the coefficient of 0.0117 on birth order is statistically significant. This finding is essentially unchanged after controlling for a variety of additional demographic variables such as race, urban location, region of residence, and parental age and education.

So as to account more flexibly for the impact of birth order and sibship size on the likelihood of obtaining a job through a sibling, I also estimate specifications in which a series of indicator variables for birth order and sibship size are included as regressors. The first four columns in the lower panel of Table 1.2 reveal a clear monotonic positive association between birth order and the probability of finding one's job through a sibling. Conditional on sibship size and relevant demographic variables, a second-born child is just 0.64 percentage points more likely than the eldest child to receive a sibling's help, whereas a seventh- or later-born child has a 6.81 percentage point higher likelihood of being helped by a sibling. By contrast, after controlling for birth order, there is only limited evidence of a systematic relationship between sibship size and the probability of obtaining a job through a sibling, although individuals from sibships with three or more members may be slightly more likely to receive help from a sibling than those belonging to sibships of size two.

[50] See appendix A.15 for detailed summary statistics regarding the role of different relatives in job search. In the sample used in this section, 4.52 percent of individuals found their most recent employer with the help of a sibling, and 3.68 percent of individuals report that this sibling also worked for their most recent employer. When tabulated by birth order, the former percentage ranges from 1.34 for first-born children to 10.35 for seventh- or later-born children, and the latter percentage ranges from 1.05 for first-born children to 8.17 for seventh- or later-born children.

Table 1.2: Relationship of Sibship Size and Birth Order to Probability of Sibling Helping Respondent Obtain Most Recent Job

	Linear Specification							
	Receive Help from Sibling				Also Have Same Employer			
Sibship Size	0.0087		0.0019	0.0014	0.0080		0.0032	0.0030
	(0.0014)	——	(0.0019)	(0.0020)	(0.0013)	——	(0.0017)	(0.0019)
Birth Order		0.0133	0.0117	0.0101		0.0109	0.0081	0.0071
	——	(0.0017)	(0.0024)	(0.0028)	——	(0.0016)	(0.0022)	(0.0026)
Demographic Controls	No	No	No	Yes	No	No	No	Yes
R^2	0.0160	0.0224	0.0227	0.0293	0.0164	0.0194	0.0204	0.0273
Families	4303	4303	4303	4303	4303	4303	4303	4303
Individuals	4973	4973	4973	4973	4973	4973	4973	4973
	Non-Linear Specification							
	Receive Help from Sibling				Also Have Same Employer			
Sibship Size 3	0.0170		0.0095	0.0088	0.0034		-0.0014	-0.0023
	(0.0069)	——	(0.0070)	(0.0070)	(0.0059)	——	(0.0062)	(0.0062)
Sibship Size 4	0.0201		0.0030	0.0028	0.0160		0.0036	0.0033
	(0.0071)	——	(0.0073)	(0.0076)	(0.0068)	——	(0.0070)	(0.0072)
Sibship Size 5	0.0394		0.0133	0.0114	0.0313		0.0114	0.0099
	(0.0094)	——	(0.0100)	(0.0103)	(0.0087)	——	(0.0088)	(0.0091)
Sibship Size 6	0.0432		0.0096	0.0068	0.0289		0.0032	0.0011
	(0.0111)	——	(0.0117)	(0.0125)	(0.0099)	——	(0.0105)	(0.0113)
Sibship Size 7+	0.0643		0.0150	0.0108	0.0541		0.0188	0.0161
	(0.0090)	——	(0.0106)	(0.0113)	(0.0085)	——	(0.0101)	(0.0107)
Birth Order 2		0.0098	0.0093	0.0064		0.0094	0.0087	0.0068
	——	(0.0054)	(0.0055)	(0.0060)	——	(0.0051)	(0.0052)	(0.0056)
Birth Order 3		0.0306	0.0275	0.0220		0.0216	0.0189	0.0152
	——	(0.0075)	(0.0081)	(0.0091)	——	(0.0065)	(0.0069)	(0.0081)
Birth Order 4		0.0535	0.0501	0.0433		0.0459	0.0393	0.0342
	——	(0.0103)	(0.0111)	(0.0122)	——	(0.0094)	(0.0101)	(0.0111)
Birth Order 5		0.0639	0.0568	0.0487		0.0572	0.0474	0.0423
	——	(0.0138)	(0.0150)	(0.0158)	——	(0.0129)	(0.0140)	(0.0147)
Birth Order 6		0.0790	0.0719	0.0656		0.0662	0.0551	0.0508
	——	(0.0188)	(0.0202)	(0.0206)	——	(0.0173)	(0.0185)	(0.0191)
Birth Order 7+		0.0881	0.0793	0.0681		0.0693	0.0535	0.0470
	——	(0.0161)	(0.0187)	(0.0207)	——	(0.0148)	(0.0173)	(0.0192)
Demographic Controls	No	No	No	Yes	No	No	No	Yes
R^2	0.0152	0.0236	0.0240	0.0303	0.0157	0.0207	0.0217	0.0283
Families	4303	4303	4303	4303	4303	4303	4303	4303
Individuals	4973	4973	4973	4973	4973	4973	4973	4973

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for gender and birth year fixed effects. Demographic controls include indicators for race, urban location, region of residence, mother's education, father's education, mother's age, father's age, and dummies for missing data on a given variable. The estimates above are based on responses to the questions on job search methods in the 1982 round of the NLSY79. The sample used here consists of respondents who have left school for the first time and are working at a full-time job when surveyed in 1982. Individuals are excluded from the analysis if they report being an only child or have missing data on birth order or sibship size.

In order to confirm that the observed positive relationship between birth order and the use of a sibling in job search is due in part to siblings working for the same employer, the second four columns of Table 1.2 present estimates of the impact of birth order and sibship size on the likelihood that an individual obtained her most recent job with the help of a sibling who was working for the employer that offered her the job. The pattern of results in the second four columns is largely similar to that in the first four columns. In the upper panel, when both sibship size and birth order are added to the regression along with relevant background variables, the coefficient on sibship size is not significantly different from zero, whereas the positive coefficient on birth order is statistically significant. In the lower panel, where indicator variables for sibship size and birth order are included as regressors, the probability of obtaining a job through a sibling appears to be more strongly related to birth order than to sibship size.

Overall, there is substantial evidence of a positive association between birth order and the use of a sibling in job search.[51] One way to interpret this finding is as follows. Consider the model of job search through social networks in Montgomery (1991). In that setup, social groups consist of either one or two members with groups of size two containing an older and a younger agent. The older agent always finds an employer without intervention by the younger agent, but the younger agent sometimes obtains a job with assistance from the older agent. The results in this section indicate that the job search patterns of later- and earlier-born children resemble the behavior of younger and older agents in Montgomery (1991). In particular, a later-born child is more likely than an earlier-born child to obtain a job through a sibling in a similar way that a younger agent unlike an older agent can find an employer through a personal contact.

This analogy can be taken further. In Montgomery (1991), an employer that hires a younger worker through an older worker learns the older agent's ability based on her performance but has imperfect information about the younger agent's productivity. Since the abilities of the two workers in each social group are correlated, the employer uses its knowledge of the older agent's performance to infer the younger agent's productivity when setting wages. Consequently, a younger worker's wage incorporates information about an older worker's ability but not vice versa. The empirical question that arises is whether a younger sibling's wage is more closely tied than an

[51] Moreover, appendix A.15 conducts a family fixed-effects analysis of the influence of birth order on job search behavior. The within-family estimates also indicate that birth order has a significant positive impact on the use of a sibling in job search.

older sibling's wage to the ability of another sibling.[52] That is, does the basic informational mechanism in Montgomery (1991) also apply to the wages of siblings? In the next section, I examine this question.

1.6.2 Sibling AFQT Impacts

This section presents the main estimates of sibling effects. Table 1.3 displays the impacts of one's own and a sibling's AFQT scores on a person's log wage. The estimated coefficients on an older sibling's AFQT score and schooling in a younger sibling's log wage are respectively greater than and less than the estimated coefficients on a younger sibling's AFQT score and schooling in an older sibling's log wage. In addition, an older sibling has a higher estimated coefficient on one's own AFQT score and a lower estimated coefficient on one's own schooling than a younger sibling. The latter pair of findings is consistent with the empirical results on employer learning in Altonji and Pierret (2001), who observe a rise in the coefficient on AFQT scores and a fall in the coefficient on schooling as workers gain labor market experience. In some specifications, the estimated coefficient on a younger sibling's AFQT score in an older sibling's log wage is negative. This finding does not necessarily suggest that a sibling's performance has a negative causal effect on a person's log wage, even though such a phenomenon is theoretically possible in a more general model where the measurement errors in the performance signals are correlated among siblings. Instead, as explained in section 1.2.2, a negative coefficient on a sibling's test score can be attributed to a positive correlation between the error terms in siblings' test scores.

I next differentiate between the two learning models in section 1.2 based on the predictions in propositions 1.2.3 and 1.2.4. Using the estimates in the fifth and sixth columns of Table 1.3, I perform a Wald test of the nonlinear hypothesis that the ratio of the coefficient on a sibling's AFQT score to that on one's own AFQT score is the same in both a younger and an older sibling's log wage. This restriction is rejected at the one percent level for both specifications. In particular, the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is significantly greater than the ratio of the coefficient on a younger

[52] A potential concern with the empirical results here is that the percentages of individuals obtaining a job through a sibling or also working at the same firm as a sibling might be too small to produce the substantial estimates of sibling effects seen in section 1.6.2. To address this issue, appendix A.13 presents a simple model of employee referrals that can generate sibling effects even if siblings work for different firms in equilibrium.

Table 1.3: Impact of Own AFQT and AFQT of Younger or Older Sibling in Labor Market on Log Wage

Older Sibling's AFQT × Younger Sibling	0.0409	0.0261	0.0304	0.0243	0.0270	0.0222
	(0.0105)	(0.0101)	(0.0107)	(0.0099)	(0.0118)	(0.0111)
Younger Sibling's AFQT × Older Sibling	0.0188	0.0027	0.0059	-0.0022	-0.0142	-0.0184
	(0.0104)	(0.0104)	(0.0102)	(0.0100)	(0.0111)	(0.0110)
Own AFQT × Younger Sibling	0.1704	0.1507	0.0997	0.0939	0.1001	0.0942
	(0.0114)	(0.0115)	(0.0115)	(0.0114)	(0.0115)	(0.0114)
Own AFQT × Older Sibling	0.2259	0.2031	0.1600	0.1495	0.1650	0.1550
	(0.0117)	(0.0117)	(0.0129)	(0.0124)	(0.0128)	(0.0123)
Own Schooling × Younger Sibling			0.0524	0.0498	0.0516	0.0493
	——	——	(0.0042)	(0.0042)	(0.0044)	(0.0043)
Own Schooling × Older Sibling			0.0468	0.0435	0.0410	0.0390
	——	——	(0.0042)	(0.0044)	(0.0043)	(0.0046)
Older Sibling's Schooling × Younger Sibling					0.0025	0.0017
	——	——	——	——	(0.0039)	(0.0039)
Younger Sibling's Schooling × Older Sibling					0.0161	0.0143
	——	——	——	——	(0.0042)	(0.0041)
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0083	0.0092
R^2	0.3278	0.3425	0.3580	0.3670	0.3595	0.3681
Families	1993	1993	1993	1993	1993	1993
Individuals	4726	4726	4726	4726	4726	4726
Sibling Pairs	7074	7074	7074	7074	7074	7074
Observations	54474	54474	54474	54474	54474	54474

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. The main estimation sample described in the text is used here.

sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage. This finding provides strong evidence against the individual learning model in proposition 1.2.3 but is consistent with the social learning model in proposition 1.2.4. An older sibling's ability seems to have a larger influence on a younger sibling's log wage than vice versa.

Table 1.4 illustrates how the main estimates vary depending on whether the members of each sibling pair are working in the same or different occupations or industries.[53] If labor market interactions are generating the observed sibling effects, then these findings might be expected to strengthen among siblings working in the same occupation or industry, because such individuals would be more likely to have come into contact with each other's employer. In the upper panel, the first, second, and third pairs of columns display results for siblings working in the same occupation, industry, and occupation or industry, respectively. An older sibling's AFQT score has a large positive estimated impact on a younger sibling's log wage, whereas a younger sibling's AFQT score has a small negative estimated impact on an older sibling's log wage. Notably, the estimated impact of an older sibling's AFQT score on a younger sibling's log wage is greater than the estimated impact of a younger sibling's own AFQT score. In the lower panel, the first, second, and third pairs of columns display results for siblings working in different occupations, industries, and occupations and industries, respectively. For both younger and older siblings, the estimated impact of a sibling's AFQT score on a person's log wage is smaller than the estimated impact of one's own AFQT score. In every specification, the restriction arising under individual learning can be rejected at least at the five percent level of significance with the deviation from the null hypothesis being consistent with social learning.[54] Nonetheless, the asymmetries between the impacts of older and younger siblings' AFQT scores on a person's log wage appear to be more pronounced when siblings work in the same as opposed to different fields.

---

[53]The 2000 Census 3-digit codes for the occupation and industry of each job are used to construct the samples for this exercise. Between the 1979 and 2000 rounds of the NLSY79, the occupation and industry of every job were entered as 1970 Census 3-digit codes. Using the crosswalks available from the U.S. Census Bureau, these fields were sequentially mapped into 1980, 1990, and 2000 Census 3-digit codes based on the occupation or industry with the most workers in a given Census year for each occupation or industry in the previous Census year.

[54]Because of the narrow definition of occupations and industries used here, siblings may be employed at the same firm even if they work in different fields, and individuals may have some contact with firms outside of their exact field. Moreover, siblings currently working in different occupations and industries may have previously worked in the same occupation or industry. Therefore, some evidence of social effects might be expected even if siblings work in different fields.

Table 1.4: Impact on Log Wage of Own AFQT and AFQT of Younger or Older Sibling Working in Same or Different Occupation or Industry

	Same Occupation		Same Industry		Either or Both	
Older Sibling's AFQT × Younger Sibling	0.0919	0.0862	0.0972	0.0901	0.0992	0.0915
	(0.0356)	(0.0335)	(0.0257)	(0.0247)	(0.0248)	(0.0230)
Younger Sibling's AFQT × Older Sibling	-0.0242	-0.0184	-0.0304	-0.0094	-0.0250	-0.0124
	(0.0332)	(0.0317)	(0.0269)	(0.0258)	(0.0247)	(0.0233)
Own AFQT × Younger Sibling	0.0548	0.0461	0.0640	0.0740	0.0529	0.0539
	(0.0334)	(0.0345)	(0.0281)	(0.0270)	(0.0253)	(0.0254)
Own AFQT × Older Sibling	0.2127	0.1953	0.1514	0.1227	0.1731	0.1491
	(0.0271)	(0.0275)	(0.0279)	(0.0270)	(0.0234)	(0.0234)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.0230	0.0265	0.0098	0.0348	0.0029	0.0066
R^2	0.4419	0.4810	0.4146	0.4554	0.4108	0.4428
Families	486	486	587	587	746	746
Individuals	1060	1060	1290	1290	1674	1674
Sibling Pairs	1202	1202	1476	1476	1968	1968
Observations	2424	2424	4040	4040	5364	5364
	Different Occupation		Different Industry		Both	
Older Sibling's AFQT × Younger Sibling	0.0257	0.0214	0.0231	0.0193	0.0213	0.0175
	(0.0118)	(0.0112)	(0.0120)	(0.0113)	(0.0118)	(0.0112)
Younger Sibling's AFQT × Older Sibling	-0.0119	-0.0157	-0.0083	-0.0131	-0.0090	-0.0138
	(0.0112)	(0.0111)	(0.0113)	(0.0112)	(0.0114)	(0.0113)
Own AFQT × Younger Sibling	0.1027	0.0973	0.1053	0.0989	0.1068	0.1006
	(0.0116)	(0.0115)	(0.0117)	(0.0115)	(0.0117)	(0.0116)
Own AFQT × Older Sibling	0.1646	0.1556	0.1681	0.1593	0.1667	0.1581
	(0.0131)	(0.0125)	(0.0132)	(0.0127)	(0.0133)	(0.0128)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.0147	0.0167	0.0384	0.0351	0.0447	0.0433
R^2	0.3573	0.3656	0.3597	0.3680	0.3584	0.3666
Families	1964	1964	1952	1952	1946	1946
Individuals	4652	4652	4623	4623	4609	4609
Sibling Pairs	6924	6924	6856	6856	6830	6830
Observations	47878	47878	46262	46262	44938	44938

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In the upper panel, the subsample in the first (second) pair of columns is composed of those observations from the main estimation sample in which both siblings work in the same 2000 Census 3-digit occupation (industry), and the subsample in the third pair of columns includes any observation belonging to either or both of the samples in the first and second pairs of columns. In the lower panel, the subsample in the first (second) pair of columns is composed of those observations from the main estimation sample in which both siblings work in different 2000 Census 3-digit occupations (industries), and the subsample in the third pair of columns includes any observation belonging to both the samples in the first and second pairs of columns. Between the 1979 and 2000 rounds of the NLSY79, the occupation and industry of each job were originally recorded as 1970 Census 3-digit codes. To construct the datasets used here, these fields are converted to 2000 Census 3-digit codes based on the crosswalks available from the U.S. Census Bureau.

Although social learning is a plausible explanation for the empirical results, there are other models that could give rise to similar patterns in the data. One possibility is that the results do reflect labor market interactions between siblings but that an informational mechanism such as employer learning is not involved. In particular, some of the existing literature on nepotism is based on a human capital framework in which individuals earn returns to familial connections. For example, Lam and Schoeni (1993) attempt to test for nepotism by comparing the coefficients on a father's and a father-in-law's schooling in a log wage equation. In the current setting, one can perform an analogous test using the coefficients on an older and a younger sibling's schooling in the fifth and sixth columns of Table 1.3.[55] If nepotistic returns to a sibling's human capital were driving the empirical results, then one might expect an older sibling's schooling to have a higher coefficient than a younger sibling's schooling in an individual's log wage equation. However, I obtain the opposite result. The coefficient on a younger sibling's schooling in an older sibling's log wage is greater than the coefficient on an older sibling's schooling in a younger sibling's log wage.[56] Thus, nepotistic connections associated with a sibling's human capital do not appear to explain the empirical results.

Another potential factor might be taste-based nepotism, which is discussed in Becker's (1971) study of discrimination. Specifically, employers may exhibit favoritism towards the younger sibling of an older employee, or an older individual may wish to use her influence within a firm to assist a younger sibling. However, such behavior does not seem to provide a satisfactory explanation for the empirical results in Table 1.3. First, it is unclear whether preferential treatment given to the older sibling of a younger worker would be associated with human capital measures such as AFQT scores and education. Second, even if such preferences were connected to human capital, there is little reason to believe that they would be positively related to a difficult-to-observe variable like the AFQT score but negatively related to an easy-to-observe variable like schooling, unless some sort of informational mechanism is involved. By contrast, the social learning model in section 1.2.4 is capable of generating both a higher coefficient on an older compared to a younger sibling's AFQT score and a lower coefficient on an older compared to a younger sibling's school-

[55] Similar results are obtained if the proposed test is instead performed using the subsamples in Table 1.4.

[56] The two-sided p-values for this test are respectively 0.0171 and 0.0210 for the specifications in the fifth and sixth columns of Table 1.3.

ing, although there also exist some parameter values for which it does not predict both of these outcomes.

In addition, several mechanisms other than labor market interactions could play a role in producing the asymmetries observed in Table 1.3. First, there might be human capital transfers between siblings, especially while siblings are living together in the parental home. For example, Zajonc (1976) notes that older children might serve as teachers for younger children within a household. Therefore, an individual's human capital may be more closely linked to the skills of an older compared to a younger sibling. Second, even if siblings do not directly transfer human capital to each other, other factors might give rise to asymmetries in the relationship of an individual's skills to the abilities of a younger and an older sibling. For example, Cunha and Heckman (2007) observe that the critical period for the development of one's cognitive skills occurs early in life, particularly during the first ten years of childhood. Because an older sibling would tend to be present for a greater portion of this critical period than a younger sibling, an older sibling may have a greater influence than a younger sibling on the formation of one's skills. Third, an older sibling might serve as a role model for a younger sibling. That is, the actions of an older sibling may provide signals of appropriate behavior for a younger sibling to follow. For example, Butcher and Case (1994) find that women raised only with brothers in their family have greater educational attainment than women with at least one sister in their family, perhaps because education is a masculine trait and women with brothers imitate such traits. In the current setting, if human capital investment is a characteristic of able individuals and older siblings are stronger role models than younger siblings, then a more able older sibling may be expected to have a greater positive effect on a person's skill acquisition than a more able younger sibling.

All three factors mentioned above—human capital transfers, asymmetries in skill formation, and role model effects—might contribute to the findings in Table 1.3, which indicate that an older sibling's AFQT score has a greater impact on a younger sibling's log wage than vice versa. Nonetheless, these mechanisms are unrelated to wage setting in particular and would likely affect other human capital measures. Specifically, if interactions among siblings prior to labor market entry were driving the empirical results, then one would expect test scores or schooling to exhibit the same asymmetric relationships as the log wage. In order to test for such an effect, I assemble a dataset consisting of one observation on each sibling pair in the main estimation sample and regress one sibling's AFQT score and schooling on the AFQT score of the other sibling in the pair.

The coefficient on the other sibling's AFQT score is estimated separately depending on whether the sibling whose AFQT score and schooling are used as dependent variables is the older or the younger sibling in the pair. Table 1.5 displays the impact of an individual's AFQT score on the AFQT score and schooling of a younger or an older sibling. Although there is some evidence of differences in the impact of a younger compared to an older sibling's AFQT score, the asymmetries observed in Table 1.5 all have the opposite direction from those in Table 1.3. In particular, the estimates in Table 1.5 indicate that, if anything, a younger sibling's AFQT score has a larger impact on an older sibling's AFQT score and schooling than vice versa. Therefore, in the absence of labor market interactions among siblings, it is difficult to explain why an older compared to a younger sibling's AFQT score has a greater impact on a person's log wage.[57]

Nonetheless, there are two issues concerning the interpretation of the results in the first two columns of Table 1.5. First, when regressing a younger sibling's AFQT score on an older sibling's AFQT score and vice versa, one is essentially performing reverse regressions. Unless the variance of an older sibling's AFQT score differs from the variance of a younger sibling's AFQT score, the coefficient on an older sibling's AFQT score in the regression of a younger on an older sibling's AFQT score must be the same as the coefficient on a younger sibling's AFQT score in the regression of an older on a younger sibling's AFQT score. Thus, the estimates in the first two columns simply suggest that the conditional variance of an older sibling's AFQT score may be slightly greater than the conditional variance of a younger sibling's AFQT score given the other control variables included in the regressions. Therefore, these specific results are largely uninformative about potential social effects among siblings.[58] Second, a possible reason for the observed difference in the conditional variance of AFQT scores between younger and older siblings might be that the variance of testing error is greater for older than for younger siblings.[59] For example,

---

[57]Nevertheless, one might argue that human capital transfers and role model effects become relevant after labor market entry, even though they do not seem to affect pre-labor market measures of skills. One piece of evidence against this possibility is that the estimated coefficient on an older sibling's schooling in a younger sibling's log wage is less than vice versa. Moreover, I provide further evidence against such factors in section 1.6.5, which conducts a series of falsification exercises.

[58]To address this first issue, appendix A.16 presents results in which the AFQT score of a middle sibling is regressed on the AFQT scores of both her youngest and her oldest sibling. In such a regression, the coefficient on the AFQT score of the youngest sibling can differ from the coefficient on the AFQT score of the oldest sibling, even if the variances of the youngest and the oldest sibling's AFQT scores are the same.

[59]In the model from section 1.2.1, testing error is represented by ω_i in equation (1.5).

Table 1.5: Relationship of Own AFQT and Schooling to AFQT of Younger or Older Sibling

	AFQT		Schooling			
Older Sibling's AFQT $\times$ Younger Sibling	0.5279	0.4120	1.0469	0.6251	0.3117	0.1334
	(0.0197)	(0.0203)	(0.0544)	(0.0552)	(0.0416)	(0.0434)
Younger Sibling's AFQT $\times$ Older Sibling	0.5640	0.4390	1.1714	0.7370	0.3859	0.2132
	(0.0185)	(0.0203)	(0.0505)	(0.0533)	(0.0385)	(0.0410)
Own AFQT	——	——	——	——	1.3927	1.1932
					(0.0395)	(0.0420)
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.0176	0.0633	0.0093	0.0147	0.0726	0.0548
R^2	0.4821	0.5288	0.2313	0.3361	0.4116	0.4565
Families	1993	1993	1993	1993	1993	1993
Individuals	4726	4726	4726	4726	4726	4726
Sibling Pairs	7074	7074	7074	7074	7074	7074

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable and fixed effects for each of the two siblings' years of birth. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. All specifications include an indicator for whether the respondent is the older or the younger sibling in a given pair. The dataset contains the first observation on every sibling pair appearing in the main estimation sample.

a larger portion of the variation in the test scores of older individuals may be attributable to factors like educational systems or cultural influences that are not fully captured by basic background variables. Such dissimilarities in the composition of test scores could contribute to asymmetries between younger and older siblings in the impact of their AFQT scores on each other's log wage.[60]

In addition, when regressing a younger sibling's schooling on an older sibling's AFQT score and vice versa, the estimates in the third and fourth columns, which do not control for one's own AFQT score, may be preferable to those in the fifth and sixth columns, in which one's own AFQT score is included as a regressor. Specifically, because an individual's AFQT score may to some extent be an endogenous function of one's own schooling, it may be difficult to interpret estimates in which an individual's own AFQT score is used as a regressor while one's own schooling is the dependent variable in the regression.[61] In any case, none of the estimates suggest that an older sibling's AFQT score has a greater impact on a younger sibling's schooling than vice versa. On the contrary, there is significant evidence of the opposite relationship in some specifications.

As a further test for sibling effects on non-wage outcomes, Table 1.6 reports the impacts of one's own and a sibling's AFQT scores on the probabilities of marriage, children, disability, and incarceration. For the analysis here, the dataset is constructed by expanding the main estimation sample to contain observations in which one or both siblings may not have valid wage data on a full-time job and by limiting the resulting sample to observations in which both siblings have information on the relevant non-wage outcomes.[62] The dependent variables are indicators for being legally married, having ever had a child, having a work disability, and currently residing in jail. The explanatory variables include one's own and a sibling's AFQT scores and schooling as

[60]To address this second issue, appendix A.18 analyzes the log wages of younger and older siblings upon reaching the same age level. As explained there, if the variance of testing error is greater for older than for younger siblings, then the resulting asymmetries between siblings in the impacts of AFQT scores on log wages would tend to have the opposite direction from those observed in the data.

[61]This issue is examined further in appendix A.20, which documents the relationship of one's AFQT score and schooling to one's own and a sibling's heights. Using height as a measure of cognitive ability that is unlikely to be affected by educational investments, there is no evidence that a younger sibling's AFQT score or schooling is more strongly related to an older sibling's height than vice versa.

[62]Including respondents without wage data in this exercise is desirable because many of the non-wage outcomes being analyzed are closely related to the likelihood of being employed. For example, incarcerated or disabled individuals are comparatively unlikely to be working.

in the fifth and sixth columns of Table 1.3.[63] For the most part, the estimated impacts of an older and a younger sibling's AFQT scores are small in size and similar to each other.[64] To assess the statistical significance of the findings, I apply the test from the fifth and sixth columns of Table 1.3 to the specifications here. Testing whether the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's outcome equation is equal to the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's outcome equation, the null hypothesis cannot be rejected for any set of coefficients as the p-values are considerably above the conventional thresholds for significance. Based on the results in Table 1.6, I do not find substantial evidence of sibling effects on non-wage outcomes.

1.6.3 Robustness Checks

The supplemental appendices report a number of robustness checks on the main estimation results. Because the evidence for sibling effects may be strongest when the age difference between siblings is greatest, appendix A.16 estimates the impacts of the AFQT scores of one's youngest and oldest siblings in the labor market on one's log wage, AFQT score, and schooling. Appendix A.17 documents the relationship between log wages and AFQT scores after modifying the empirical specification to allow for social interactions among more than two siblings in a family. In order to tighten the comparison between the log wages of younger and older siblings, appendix A.18 analyzes the log wages of younger and older siblings upon reaching the same age level. Appendix A.19 seeks to account more adequately for the causal effect of schooling on AFQT scores and presents results that control for schooling at AFQT administration. In addition, the analysis there examines the determinants of sibling correlations in human capital measures. Overall, I continue to find significant evidence of sibling effects on log wages after performing each of these robustness checks.

[63] In particular, the regressors in the upper and lower panels of Table 1.6 are respectively the same as those in the fifth and sixth columns of Table 1.3.

[64] Nevertheless, some differences between the coefficients on an older and a younger sibling's AFQT scores might be expected, because these variables are seen to have asymmetric impacts on the log wage, and labor income may affect the decision to get married, raise children, seek medical care, or commit a crime.

Table 1.6: Impact of Own AFQT and AFQT of Younger or Older Sibling on Non-Wage Outcomes

	Married	Has Kids	Disabled	In Jail
Older Sibling's AFQT × Younger Sibling	0.0046	-0.0186	0.0004	-0.0031
	(0.0090)	(0.0086)	(0.0038)	(0.0014)
Younger Sibling's AFQT × Older Sibling	0.0043	-0.0172	0.0040	-0.0013
	(0.0099)	(0.0097)	(0.0045)	(0.0013)
Own AFQT × Younger Sibling	0.0364	-0.0069	-0.0181	-0.0044
	(0.0106)	(0.0102)	(0.0043)	(0.0016)
Own AFQT × Older Sibling	0.0656	0.0016	-0.0308	-0.0043
	(0.0107)	(0.0114)	(0.0052)	(0.0016)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes
Family Background Controls	No	No	No	No
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.8387	0.5831	0.6579	0.4433
R^2	0.1352	0.2416	0.0298	0.0242
Families	2169	2169	2169	2169
Individuals	5168	5168	5168	5168
Sibling Pairs	7988	7988	7988	7988
Observations	119708	119708	119708	119708
	Married	Has Kids	Disabled	In Jail
Older Sibling's AFQT × Younger Sibling	0.0074	-0.0168	-0.0005	-0.0033
	(0.0091)	(0.0087)	(0.0038)	(0.0014)
Younger Sibling's AFQT × Older Sibling	0.0095	-0.0166	0.0041	-0.0017
	(0.0098)	(0.0096)	(0.0044)	(0.0013)
Own AFQT × Younger Sibling	0.0404	-0.0049	-0.0186	-0.0047
	(0.0106)	(0.0102)	(0.0042)	(0.0016)
Own AFQT × Older Sibling	0.0662	0.0046	-0.0301	-0.0044
	(0.0108)	(0.0115)	(0.0052)	(0.0015)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes
Family Background Controls	Yes	Yes	Yes	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.8919	0.5113	0.4849	0.5541
R^2	0.1477	0.2507	0.0371	0.0317
Families	2169	2169	2169	2169
Individuals	5168	5168	5168	5168
Sibling Pairs	7988	7988	7988	7988
Observations	119708	119708	119708	119708

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's outcome is used as the dependent variable for a given pair. The dataset used here is constructed as follows. First, the main estimation sample is expanded to include observations on sibling pairs in which one or both members may not have valid wage data on a full-time job. Second, the resulting sample is restricted to observations on sibling pairs in which both members have non-missing data on marital status, presence of children, health restrictions, and residence type. In the third column, respondents are classified as disabled if they have a health condition that prevents them from working or that limits the kind or amount of work that they can do. In the fourth column, respondents are classified as being in jail if they are residing in a correctional institution when interviewed.

1.6.4 Additional Predictions

This section tests some further implications of the individual and social learning models. In appendix A.6, I characterize the coefficients obtained from the regression of one's log wage on one's own and a sibling's schooling. Under social learning, one's own and a sibling's schooling are both easy-to-observe variables. Hence, the coefficients on one's own and an older sibling's schooling in a younger sibling's log wage should be the same as the corresponding coefficients on one's own and a younger sibling's schooling in an older sibling's log wage.[65] Under individual learning, one's own schooling is an easy-to-observe variable, but a sibling's schooling is a hard-to-observe variable. Thus, the coefficients on one's own and an older sibling's schooling in a younger sibling's log wage can differ from the corresponding coefficients on one's own and a younger sibling's schooling in an older sibling's log wage. The directions of the predicted differences between these coefficients will depend on the relative values of the sibling correlations in ability and in schooling. For example, if ability is more highly correlated among siblings than schooling, then an older sibling should have a lower coefficient on one's own schooling than a younger sibling, and the coefficient on a younger sibling's schooling in an older sibling's log wage should be higher than vice versa.[66]

Table 1.7 displays the estimates from regressing one's log wage on one's own and a sibling's schooling. In the first two columns, one's own schooling, but not a sibling's, is included as a regressor. Because one's own schooling is an easy-to-observe variable both under individual and under social learning, the coefficient on one's own schooling should be the same in a younger and an older sibling's log wage. Consistent with both learning models, the hypothesis that the coefficient on one's own schooling is the same for a younger and an older sibling cannot be rejected at conventional levels.[67] In the second two columns, a sibling's schooling is added to the regressions. Consistent with social learning, one can reject neither the hypothesis that the coefficient on

[65]This result is a counterpart to the first prediction of the employer learning model in Farber and Gibbons (1996), who show that the coefficient on schooling in a wage regression should not change with labor market experience, provided that the other regressors are also easily observable to employers.

[66]These results are analogous to the implications of the statistical discrimination model in Altonji and Pierret (2001), who show that the coefficients on test scores and schooling in a log wage regression should respectively increase and decrease with labor market experience, especially if these variables are the only two regressors.

[67]The two-sided p-values for this test are 0.1794 in the first column and 0.4992 in the second column.

one's own schooling is the same in a younger as in an older sibling's log wage nor the hypothesis that the coefficient on an older sibling's schooling in a younger sibling's log wage is the same as vice versa.[68] In addition, as explained in section 1.2.2, the significantly positive coefficient on a sibling's schooling indicates that the sibling correlation in ability is greater than that in schooling, especially assuming that schooling is not measured with substantial error or that the measurement error in schooling is highly correlated between siblings.[69]

Another difference between the implications of the individual and the social learning model concerns the variance of the change in log wage residuals between successive age levels.[70] In appendix A.7, I characterize the change between two age levels in the residual from the regression of one's log wage on one's own and a sibling's schooling. Under individual learning, the variance of the change in log wage residuals between two given ages should be the same for both a younger and an older sibling, given that individuals of varying parities are treated symmetrically in the model. Under social learning, this quantity should be smaller for a younger than for an older sibling, and the absolute difference in this quantity between a younger and an older sibling should be increasing in the size of the age gap between them. Intuitively, a younger sibling at a given age has more signals of an older sibling's performance than an older sibling at that age has of a younger sibling's performance; so that, the market forms more precise beliefs about the productivity of a younger sibling at a given age than an older sibling at that age. Hence, additional signals about one's own or a sibling's performance should have less of an impact on the market's beliefs about a younger than an older sibling's productivity. Because the wage is assumed to be equal to the conditional expectation of productivity given the market's information, the variance of the change in log wage residuals should be smaller for a younger than for an older sibling, and this effect should be larger in size when the age gap between the two siblings is wider.

[68] The respective two-sided p-values in the third and fourth columns are 0.5208 and 0.6950 for the test of equality between the coefficients on one's own schooling and 0.4089 and 0.5171 for the test of equality between the coefficients on a sibling's schooling.

[69] In appendix A.22, I attempt to account for the effect of measurement error in the schooling variable by using a variant of the instrumental-variables procedure in Ashenfelter and Krueger (1994) and Bronars and Oettinger (2006). The results there suggest that the positive coefficient observed on a sibling's schooling in the regression of one's log wage or test score on one's own and a sibling's schooling is unlikely to be merely an artifact of measurement error.

[70] Kahn (2009) develops a methodology for detecting asymmetric information between employers based on the variance of log wage change residuals.

Table 1.7: Impact of Own Schooling and Schooling of Younger or Older Sibling in Labor Market on Log Wage

Own Schooling × Younger Sibling	0.0769 (0.0037)	0.0671 (0.0040)	0.0701 (0.0042)	0.0637 (0.0042)
Own Schooling × Older Sibling	0.0829 (0.0038)	0.0705 (0.0041)	0.0741 (0.0043)	0.0660 (0.0044)
Older Sibling's Schooling × Younger Sibling	——	——	0.0141 (0.0037)	0.0095 (0.0036)
Younger Sibling's Schooling × Older Sibling	——	——	0.0190 (0.0041)	0.0130 (0.0039)
Family Background Controls	No	Yes	No	Yes
test for equality between coefficients on own schooling (p-value)	0.1794	0.4992	0.5208	0.6950
test for equality between coefficients on sibling's schooling (p-value)	——	——	0.4089	0.5171
R^2	0.3227	0.3403	0.3266	0.3419
Families	1993	1993	1993	1993
Individuals	4726	4726	4726	4726
Sibling Pairs	7074	7074	7074	7074
Observations	54474	54474	54474	54474

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. The main estimation sample described in the text is used here.

In order to test for differences between younger and older siblings in the variance of the change in log wage residuals between successive age levels, I construct a sample of sibling pairs in which each wage observation on an older sibling at a given age is matched with a wage observation on her younger sibling at that age. Specifically, I identify, for a given age level, those individuals in the NLSY79 who, when interviewed at that age, have left school for the first time, have a wage observation on a full-time job, and have a non-twin sibling also satisfying the preceding two criteria at that age. If a pair of siblings meets these conditions at no less than two different age levels, then the sample includes one observation on the sibling pair for every age at which these conditions are met.[71] In short, the resulting dataset is similar to the main estimation sample, except that wage observations on older and younger siblings are matched at a given age level instead of in a given survey year. The variance of the change in log wage residuals is estimated as follows. First, I calculate the residuals from regressing the log wage of the first sibling in a pair on the schooling and background variables of both siblings in the pair.[72] Second, I chronologically sort the log wage observations on every sibling pair, subtract the log wage residual for each observation on a sibling pair from the log wage residual for the next observation on the sibling pair, and square the result. Third, I regress the squared change in log wage residuals on a dummy variable equal to one if the first sibling in a pair is younger than the second sibling in the pair and equal to zero otherwise.

Table 1.8 presents the results from this procedure. Consistent with social learning, the point estimate for the mean squared change in log wage residuals between successive age levels is smaller for younger than for older siblings. However, this effect is not statistically significant when using the full sample, perhaps because the mean age gap between younger and older siblings is not very wide. To examine this issue further, I divide the sample into terciles based on the age difference between the siblings in a pair and compute a distinct set of estimates for each of the three subsamples.[73] Although one cannot reject the hypothesis that the mean squared change in log wage

[71] Because the social learning model properly applies to the period after both siblings in a pair have entered the labor market, I exclude an observation on a pair of siblings at a given age if one sibling has not yet left school for the first time when the other sibling is surveyed at that age.

[72] The regressions also control for a third-order bivariate polynomial in the two siblings' ages as well as the interaction of both siblings' schooling and background variables with a third-order polynomial in the first sibling's age. The former control is needed because the constant term in the regression can vary with one's own age under individual learning and with both one's own and a sibling's ages under social learning. The latter control is needed because the coefficients on schooling and background variables can depend on one's own age under individual learning.

[73] Note that the social learning model predicts that the absolute difference between a younger and an older sibling

residuals is the same for younger and older siblings from the first or second tercile, this quantity is significantly lower for younger than for older siblings from the third tercile.[74] Hence, the data appear to support the prediction of the social learning model for the variance of the change in log wage residuals, provided that the age gap between younger and older siblings is sufficiently large.

1.6.5 Falsification Exercises

In this section, I conduct three falsification exercises to strengthen the argument that labor market interactions in general and social learning in particular are responsible for the sibling effects identified in section 1.6.2. The analysis up to now has focused on the impacts of siblings who are working at the same time. Nonetheless, a pertinent question is whether siblings who have not had substantial labor market experience seem to influence a person's log wage. Specifically, if the AFQT scores of older and younger siblings with little connection to the labor market have asymmetric impacts on log wages, then it would call into question the interpretation of the results in Table 1.3 as representing labor market interactions among siblings. By contrast, a failure to detect such effects using the AFQT scores of siblings without much work experience would lend support to a labor market explanation for those results.

I now perform such a comparison. To construct a sample of working individuals who have a sibling with low work experience, I identify in each year those pairs of interviewed siblings in the NLSY79 for which one member is employed at a full-time job after initially leaving school and the other member has never spent a year primarily working.[75] The resulting sample of sibling pairs is divided into two groups depending on whether the member who has never been primarily working is older or younger than the member who has left school for the first time and is thereafter employed at a full-time job. Finally, using these two groups of sibling pairs, the log wage of a younger (resp. an older) sibling in the labor market is regressed on her own AFQT score as well as that of her inexperienced older (resp. younger) sibling.

---

in the variance of the change in log wage residuals should be increasing in the size of the age gap between them.

[74]The two-sided p-values for the test of equality between younger and older siblings from the third tercile are 0.0198 in the upper panel and 0.0206 in the lower panel.

[75]As in Farber and Gibbons (1996), a respondent is defined as primarily working if she has worked in at least half the weeks since her last interview for 30 or more hours per week on average during the working weeks.

Table 1.8: Difference Between Younger and Older Siblings in Mean Squared Change in Log Wage Residuals Between Successive Age Levels

	Entire	Terciles of Absolute Difference in Sibling Ages		
	Sample	First	Second	Third
Second-Stage Regression				
Younger Sibling	-0.0097 (0.0065)	-0.0027 (0.0108)	-0.0041 (0.0096)	-0.0268 (0.0115)
Constant	0.1250 (0.0072)	0.1179 (0.0111)	0.1258 (0.0110)	0.1351 (0.0120)
R^2	0.0009	0.0013	0.0009	0.0022
First-Stage Regression				
Basic Demographic Variables	Yes	Yes	Yes	Yes
Family Background Controls	No	No	No	No
R^2	0.3129	0.3256	0.3155	0.3108
	Entire	Terciles of Absolute Difference in Sibling Ages		
	Sample	First	Second	Third
Second-Stage Regression				
Younger Sibling	-0.0095 (0.0065)	-0.0030 (0.0109)	-0.0032 (0.0094)	-0.0264 (0.0114)
Constant	0.1247 (0.0072)	0.1178 (0.0111)	0.1248 (0.0109)	0.1341 (0.0118)
R^2	0.0009	0.0013	0.0009	0.0021
First-Stage Regression				
Basic Demographic Variables	Yes	Yes	Yes	Yes
Family Background Controls	Yes	Yes	Yes	Yes
R^2	0.3315	0.3565	0.3501	0.3367
Own Mean Age Across All Log Wage Observations	29.69	28.79	30.20	30.15
Mean Change in Age Between Consecutive Log Wage Observations	2.260	2.283	2.231	2.263
Mean Absolute Difference Between Own and Sibling Age	2.459	1.271	2.182	4.033
Families	1677	832	739	725
Individuals	3967	1878	1636	1688
Sibling Pairs	5754	2082	1730	1942
Observations	35780	12674	11524	11582

Note: The estimates above are based on a sample of log wage observations on younger and older siblings at the same ages. For a given age level, the dataset comprises those individuals in the NLSY79 who, when interviewed at that age, have left school for the first time, have non-missing data on their AFQT score and schooling, have a valid wage observation on a full-time job, have non-missing sibling data including birth order and sibship size, and have a non-twin sibling also satisfying these criteria. If a pair of siblings meets these conditions at no less than two different age levels, then the sample includes one observation on the sibling pair for every age at which these conditions are met. Otherwise, the sibling pair is dropped from the analysis. An observation on a pair of siblings at a given age level is excluded if one sibling has not yet left school for the first time when the other sibling is interviewed at that age. In the upper panel, the log wage residuals are obtained from a first-stage regression of the log wage for the first member of a sibling pair on the following variables: the highest grade completed by each sibling in the pair; dummies for the race, gender, region of residence, and urban location of each sibling; indicators for missing data; interactions of the aforementioned variables with a third-order polynomial in the age of the first sibling; and a third-order bivariate polynomial in the ages of the two siblings. In the lower panel, the first-stage regression used to obtain the log wage residuals also controls for the following: sibship size, mother's education, father's education, mother's age, father's age, each of the two siblings' birth orders, and interactions of these variables with a third-order polynomial in the first sibling's age. The squared change in log wage residuals is calculated by chronologically sorting the log wage observations on every sibling pair, subtracting the log wage residual for each observation on a sibling pair from the log wage residual for the next observation on the sibling pair, and squaring the result. In the second-stage regressions, the squared change in log wage residuals is regressed on a dummy variable equal to one if the first member of a sibling pair is younger than the second member and equal to zero otherwise. A quartic time trend with base year 1993 is also included in the first- and second-stage regressions. The standard errors in parentheses are computed using the nonparametric bootstrap with 10,000 replications. Each bootstrap sample is formed by sampling families with replacement from the original dataset, and the two-stage procedure described here is performed on the observations in each sample. In the second through fourth columns, the dataset is divided into terciles depending on the age difference between the siblings in a pair, and a separate set of estimates is computed for each of the three subsamples.

Table 1.9 presents the results from these regressions. In all six specifications, the coefficient on an older sibling's AFQT score in a younger sibling's log wage is insignificantly lower than the coefficient on a younger sibling's AFQT score in an older sibling's log wage. After controlling for one's own and a sibling's schooling, the coefficients on an older and a younger sibling's AFQT scores are both negative, although not significantly so. Moreover, in no specification does one find a significant difference between the coefficients on one's own AFQT score in the log wages of older and younger siblings.[76] Using the estimates in the fifth and sixth columns to test the restriction implied by the individual learning model in proposition 1.2.3, one cannot reject the null hypothesis that the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is equal to the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage.[77] In fact, the former ratio is insignificantly less than the latter ratio, which is contrary to the predictions of the social learning model in proposition 1.2.4. Therefore, as would be expected if the asymmetric impacts of a younger and an older sibling's AFQT scores in Table 1.3 were generated by labor market interactions, I find no evidence of such differences in Table 1.9 when using the AFQT scores of younger and older siblings without substantial labor market experience.

Nonetheless, one concern with this conclusion is that the two groups of siblings being analyzed in Table 1.9 may be dissimilar to each other. For example, if the sibling correlation in ability is sufficiently lower in families with an inexperienced older sibling than in those with an inexperienced younger sibling, then such a difference could conceal an underlying positive effect that older relative to younger siblings outside the labor force may be having on a person's log wage. However, one can address this issue to some extent by using AFQT scores as a measure of ability and comparing the sibling correlations in AFQT scores. Specifically, using a dataset containing one observation on every sibling pair included in the sample from Table 1.9, I regress the AFQT score of any individual who is working full-time after having left school on the AFQT score of her sibling who has never been classified as primarily working. The coefficients on a sibling's AFQT score are estimated separately depending on whether the inexperienced sibling is older or younger

[76]The estimated coefficients in a younger sibling's log wage are somewhat more imprecise than those in an older sibling's log wage, because the sample of older siblings with an inexperienced younger sibling is larger than the sample of younger siblings with an inexperienced older sibling.

[77]The two-sided p-values for this test are 0.6771 in the fifth column and 0.8837 in the sixth column.

Table 1.9: Impact on Log Wage of Own AFQT and AFQT of Younger or Older Sibling Not Yet Primarily Working

Older Sibling's AFQT × Younger Sibling	0.0173	0.0068	0.0056	0.0048	-0.0184	-0.0105
	(0.0253)	(0.0238)	(0.0245)	(0.0234)	(0.0322)	(0.0280)
Younger Sibling's AFQT × Older Sibling	0.0232	0.0129	0.0141	0.0091	-0.0048	-0.0070
	(0.0134)	(0.0137)	(0.0132)	(0.0134)	(0.0130)	(0.0131)
Own AFQT × Younger Sibling	0.1572	0.1346	0.1070	0.0925	0.1084	0.0938
	(0.0271)	(0.0253)	(0.0269)	(0.0250)	(0.0269)	(0.0249)
Own AFQT × Older Sibling	0.1439	0.1297	0.1071	0.0996	0.1115	0.1046
	(0.0147)	(0.0145)	(0.0157)	(0.0151)	(0.0156)	(0.0152)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.6771	0.8837
R^2	0.2573	0.2805	0.2708	0.2915	0.2752	0.2951
Families	1528	1528	1528	1528	1528	1528
Individuals	2175	2175	2175	2175	2175	2175
Sibling Pairs	2670	2670	2670	2670	2670	2670
Observations	9596	9596	9596	9596	9596	9596

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. For a given survey year, the sample used here consists of those individuals in the NLSY79 who have left school for the first time, have non-missing data on their AFQT score and schooling, have a valid wage observation on a full-time job, have non-missing sibling data including birth order and sibship size, and have a non-twin sibling who has not yet spent a year primarily working. An interviewed sibling is classified as primarily working if she has worked in at least half the weeks since the last interview for an average of at least 30 hours per week during the working weeks. In every survey year, any respondent satisfying the criteria listed above is paired with each of her siblings who has not yet spent a year primarily working, and the resulting sample of sibling pairs is divided into two groups based on whether the respondent is older or younger than the sibling who has not yet been primarily working. The analysis excludes any siblings whose first year spent primarily working cannot be accurately determined because they have a positive number of weeks unaccounted for in the work history data.

than the individual whose AFQT score is used as a dependent variable. Testing for differences between the coefficients on an older and a younger sibling's AFQT scores, there is insufficient evidence to reject the null hypothesis that these two coefficients are equal.[78] Therefore, there is little indication that differences across families in the correlation among siblings' abilities are disguising significant asymmetries between the impacts of siblings' AFQT scores on the log wage in Table 1.9.

I now conduct a second round of falsification exercises, in which the AFQT scores and log wages of spatially isolated siblings are analyzed. First, I compare younger and older siblings who were residing in different geographic regions towards the beginning of their careers. If labor market interactions among siblings are generating the disparate impacts of a younger and an older sibling's AFQT scores on the log wage in Table 1.3, then such asymmetries should largely be absent when studying younger and older siblings who have worked in different regional labor markets since early in their careers, especially assuming that the observed asymmetries are driven primarily by siblings working for the same employer or that workers and firms in one geographic region have little contact with individuals residing in another region. Second, I examine pairs of siblings who start their careers working in the same region but later become geographically separated. If employers are using their knowledge of one sibling's performance to infer another sibling's productivity when setting wages, then the impact of a sibling's AFQT score on a person's log wage might decrease relative to the impact of one's own AFQT score when either of two siblings initially residing in the same geographic region moves to a different region.

In order to perform the first of these two exercises, it is important to identify a sample of siblings who are living apart from close to the outset of their careers. Otherwise, if a pair of siblings becomes separated only later in life, then it may be possible for some information on one sibling's performance to get incorporated into the other sibling's wage if the two siblings are initially working together and for such information to be transmitted between employers when either sibling changes jobs if the wage is publicly observable. Therefore, I assemble a dataset composed only of observations on those pairs of siblings who reside in different geographic regions

[78]The two-sided p-values for this test are 0.7387 when the specification controls only for basic demographic attributes as in the odd-numbered columns of Table 1.9 and 0.5450 when the specification also contains relevant family background variables as in the even-numbered columns of Table 1.9.

from each other during the first survey year when living in one's own dwelling unit.[79] Using this subset of the main estimation sample, I replicate the analysis in Table 1.3, regressing a younger sibling's log wage on an older sibling's AFQT score and vice versa.

The results from these regressions are reported in Table 1.10. Out of six specifications, there are three cases in which the coefficient on an older sibling's AFQT score is insignificantly greater than that on a younger sibling's AFQT score and three cases in which the coefficient on a younger sibling's AFQT score is insignificantly greater than that on an older sibling's AFQT score. Using the estimates in the last two columns to test the restriction imposed by the individual learning model in proposition 1.2.3, I fail to reject the null hypothesis that the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in the log wage of a younger sibling is equal to the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in the log wage of an older sibling.[80] Moreover, counter to the implications of the social learning model in proposition 1.2.4, the former ratio is insignificantly smaller than the latter ratio. Overall, I am unable to find significant evidence of differences between the impacts of the AFQT scores of younger and older siblings who were residing in different regions during the early stages of their careers. This is another result consistent with the interpretation of the relevant asymmetries in Table 1.3 as representing labor market interactions between younger and older siblings.

It appears to be possible to take the analysis of spatially isolated siblings at least one step further. This final exercise studies how the coefficient on a single sibling's AFQT score changes over time when two siblings initially living in the same region become geographically separated. To motivate the empirical findings, appendix A.12 presents some basic comparative statics results describing how the coefficients on one's own and a sibling's test scores are predicted to change in a log wage regression when an individual gains additional signals about her own or her sibling's performance. In particular, consider the ratio of the coefficient on a sibling's test score to that on one's own test score in the regression of an individual's log wage on both her own and her sibling's test scores and schooling levels. On the one hand, if learning is individual as in section

[79]The regions referred to above are the four Census geographic regions of the United States: Northeast, Midwest, South, and West.

[80]The two-sided p-values for this test are 0.7785 in the fifth column and 0.6727 in the sixth column.

Table 1.10: Impact on Log Wage of Own AFQT and AFQT of Younger or Older Sibling Residing in Different Geographic Region When First Living on Own

Older Sibling's AFQT $\times$ Younger Sibling	0.0079	-0.0144	-0.0141	-0.0254	-0.0371	-0.0503
	(0.0364)	(0.0291)	(0.0379)	(0.0301)	(0.0418)	(0.0335)
Younger Sibling's AFQT $\times$ Older Sibling	0.0011	-0.0320	-0.0120	-0.0296	-0.0280	-0.0373
	(0.0285)	(0.0290)	(0.0280)	(0.0287)	(0.0296)	(0.0310)
Own AFQT $\times$ Younger Sibling	0.1934	0.1583	0.1346	0.1237	0.1379	0.1298
	(0.0372)	(0.0351)	(0.0391)	(0.0349)	(0.0388)	(0.0347)
Own AFQT $\times$ Older Sibling	0.2349	0.1996	0.1468	0.1378	0.1479	0.1384
	(0.0351)	(0.0298)	(0.0380)	(0.0339)	(0.0381)	(0.0338)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.7785	0.6727
R^2	0.3119	0.3901	0.3501	0.4120	0.3530	0.4140
Families	271	271	271	271	271	271
Individuals	641	641	641	641	641	641
Sibling Pairs	746	746	746	746	746	746
Observations	5698	5698	5698	5698	5698	5698

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. The dataset is constructed from the main estimation sample by identifying each individual's region of residence in the first survey year in which she is living in her own dwelling unit. The estimates are based on those pairs of siblings who each reside in a different region of the United States when first living on one's own. A sibling pair is excluded from the analysis if either sibling had resided outside the four Census geographic regions of the United States as of the first survey year when living on one's own or if either sibling's first survey year when living on one's own could not be accurately determined because of missing data on type of residence.

1.2.3, then this ratio will remain constant, irrespective of the number of productivity signals that a person or her sibling has acquired. Hence, when comparing the log wages of two siblings who are initially living together but who later become spatially separated, this ratio would be expected to stay the same from before to after separation if employer learning is indeed individual. On the other hand, if learning is social as in section 1.2.4, then this ratio is typically decreasing in the number of signals about one's own performance and increasing in the number of signals about a sibling's performance. Thus, there is some reason to believe that this ratio might decrease if either of two siblings currently living in the same region moves to a different region.

The analysis of the log wages of siblings before and after being separated is conducted as follows. First, because this exercise does not involve comparing younger and older siblings, I expand the main estimation sample to include observations on pairs of siblings having the same year and month of birth.[81] Second, I identify those sibling pairs for which there exists a successive pair of survey years such that the two siblings are living in the same geographic region in the first year but not in the second year. Third, whenever such a pair of years is located for a particular sibling pair, the observation on the sibling pair in the first of these years is added to a sample representing siblings before separation, and the observation on the sibling pair in the second of these years is added to a sample representing siblings after separation. Finally, the log wages of the first member of each sibling pair are regressed on her own AFQT score and that of the second member of the sibling pair. The coefficients are estimated separately based on whether the dependent variable is the log wage before or after the siblings are residing in different regions.

The results from this procedure are presented in Table 1.11. Even though the estimates are somewhat imprecise because of the comparatively small sample sizes involved, the basic pattern of results appears to be consistent with a model of social learning in which a person's log wage incorporates information on the performance of a sibling living sufficiently nearby. Between the last survey year in which two siblings are living in the same region and the first survey year in which they are living in different regions, the point estimate for the coefficient on a sibling's AFQT score falls, and the point estimate for the coefficient on one's own AFQT score rises. Computing distinct sets of estimates based on whether the first member of a sibling pair moves to a different

[81] In addition, since this exercise does not depend on differences in birth order between siblings, the sample used here also includes sibling pairs in which one or both members may be missing data on their number of older siblings.

job or stays with the same employer between the two survey years, the decline in the coefficient on a sibling's AFQT score appears to be more pronounced for individuals changing their jobs than for those remaining with their original employer, although the estimates become increasingly imprecise as one segments the sample. This finding is sensible because the current employer of a person changing jobs compared to the employer of a person staying at the same job may be less likely to have come into direct contact with the person's sibling during the period when the two siblings were residing in the same region as each other.

In order to assess the statistical significance of the pattern of results in Table 1.11, I test whether the ratio of the coefficient on a sibling's AFQT score to that on one's own AFQT score remains constant before and after the siblings in each pair become separated. For the full sample, the restriction implied by the individual learning model can be rejected at the five percent level of significance.[82] Specifically, I find that the ratio of the coefficient on a sibling's AFQT score to that on one's own AFQT score decreases significantly when two siblings residing in the same region as each other become geographically separated. When the sample is partitioned into two groups depending on whether or not the first member of a sibling pair changes employers between the two survey years, there is some suggestion of a decrease in this ratio from before to after separation in both subsamples, although the null hypothesis cannot be rejected at conventional levels in either subsample by itself.

Overall, the empirical results for the full sample in Table 1.11 lend further support to the position that employer learning is partly social as opposed to purely individual. Moreover, these results seem to provide additional evidence against the argument that the main findings in this paper are driven by human capital transfers or role model effects among siblings in the labor market. In particular, the latter two mechanisms would have some difficulty explaining why the coefficient on a sibling's AFQT score in a log wage regression decreases relative to the coefficient on one's own AFQT score when a person becomes separated from her sibling, unless human capital acquired from a sibling rapidly deteriorates during this period, or lessons learned from a sibling are suddenly forgotten at this time. That is, the findings here appear to be easiest to understand within a framework in which employers have access to information on a sibling working sufficiently nearby.

[82]The two-sided p-values for this test are 0.0431 in the first column and 0.0351 in the second column.

Table 1.11: Impact of Own and Sibling's AFQT on Log Wage Immediately Before and After Siblings Reside in Different Geographic Regions

	Entire Sample		Job Change		No Job Change	
Sibling's AFQT×Before Separated	0.0285	0.0258	0.0245	0.0400	0.0213	0.0203
	(0.0284)	(0.0281)	(0.0486)	(0.0490)	(0.0325)	(0.0355)
Sibling's AFQT×After Separated	-0.0197	-0.0320	-0.0570	-0.0463	0.0032	0.0141
	(0.0298)	(0.0294)	(0.0516)	(0.0473)	(0.0314)	(0.0331)
Own AFQT×Before Separated	0.0929	0.0930	0.0792	0.0800	0.0976	0.0951
	(0.0257)	(0.0247)	(0.0399)	(0.0401)	(0.0373)	(0.0359)
Own AFQT×After Separated	0.1223	0.1060	0.1059	0.0629	0.1171	0.1116
	(0.0277)	(0.0286)	(0.0471)	(0.0481)	(0.0331)	(0.0329)
Own Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.0431	0.0351	0.1468	0.1888	0.3937	0.7159
Families	263	263	203	203	220	220
Individuals	598	598	279	279	344	344
Sibling Pairs	692	692	329	329	380	380
Observations	1480	1480	680	680	800	800

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the dependent variable is the log wage observation before or after the siblings are separated. To construct the dataset used here, the main estimation sample is first expanded to include pairs of siblings born in the same year and month as well as sibling pairs in which one or both members may be missing data on their number of older siblings. This intermediate sample is then used to identify those sibling pairs for which there exists a consecutive pair of survey years such that the two siblings are living in the same Census geographic region of the United States in the first year but not in the second year. If such a pair of years exists for a given sibling pair, then the observations on the sibling pair for the first and second years are included in the samples of sibling pairs before and after being separated, respectively. If there is more than one such pair of years for the sibling pair, then all such pairs of years are used in the analysis. The dataset excludes any sibling pair in which either member is recorded as residing in a region other than one of the four Census geographic regions of the United States. A sibling pair is included in the job-change sample if there is a change between the two years in the CPS job of the sibling whose wage is used as the dependent variable for the pair. Otherwise, the sibling pair is added to the no-change sample.

1.7 Conclusion

This paper has constructed and implemented a test for statistical nepotism in the labor market. Embedding a sibling model into an employer learning framework, I presented a theory of labor markets with symmetric but imperfect information among employers in which workers belong to disjoint social groups and workers in the same group have similar attributes. This setup was used to study wage determination under two alternative assumptions about employers' formation of beliefs: individual learning and social learning. Under the former specification, a worker's wage incorporates information only about her own schooling and performance, whereas under the latter specification, a worker's wage can also include information about the schooling and performance of her personal contacts. Using data on the AFQT scores of siblings from the NLSY79, I applied this setup to examine whether a worker's wage contains a component based on a sibling's performance. If learning is social, then the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage should typically be greater than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage. However, if learning is individual, then these two ratios should be the same.

The empirical results provided strong support for the central prediction of the social learning model. Moreover, I presented a large body of evidence to substantiate the claim that the findings are best understood as representing social effects among siblings in employer learning. First, I documented a robust positive correlation between birth order and the probability of obtaining a job through a sibling. Second, I found that an individual's AFQT score shows no evidence of being more closely related to an older sibling's AFQT score than to a younger sibling's AFQT score. Third, there was also no evidence that an older sibling's AFQT score has a greater impact on a younger sibling's schooling than vice versa. Fourth, although an older sibling's AFQT score has a larger impact on a younger sibling's log wage than vice versa, the same regressions indicated that an older sibling's schooling has a smaller impact on a younger sibling's log wage than vice versa. Fifth, the AFQT scores of older and younger siblings without substantial labor market experience were shown to have similar effects on a person's log wage. Sixth, I found no discernable difference between the impacts of the AFQT scores of older and younger siblings who were residing in different geographic regions during the early stages of their careers. Seventh, there was some evidence of a decrease in the impact of a sibling's AFQT score when either of two siblings initially living in the same region moves to a different region.

A possible area for extending the investigation of statistical nepotism in the current paper would be to examine the schooling decisions or occupational choices of siblings. An analysis of schooling decisions might be interesting because sibling effects in employer learning could potentially contribute to the negative correlation between birth order and educational attainment documented by Behrman and Taubman (1986) and Black et al. (2005). Specifically, if an older sibling's performance provides information to employers about a younger sibling's ability, then the signaling returns to schooling may be lower for younger than for older siblings, leading younger siblings to invest less in education than their older counterparts. Alternatively, an analysis of occupational choices among siblings could help determine whether social learning can increase the efficiency of labor markets by improving the quality of information available about a worker's suitability for a given type of job.

Chapter 2

The Impact of Macroeconomic Conditions in Childhood on Adult Labor Market Outcomes

2.1 Introduction

The process of skill formation in childhood is an important area of research in labor economics. An understanding of the effects of business cycles on child development can be useful to policy-makers when designing and targeting economic stimulus plans or health care programs. This paper investigates how macroeconomic conditions during one's formative years affect one's labor market performance in adulthood. The analysis proceeds in three stages. First, a large sample from the Census is used to document the relationship of state and national unemployment rates during childhood to schooling, employment, and income as an adult. Second, a matched sample of parents and children from the Panel Study of Income Dynamics (PSID) is constructed in order to assess whether the observed impact of the macroeconomy can be attributed to selection over the business cycle in the background characteristics of parents raising children. Third, detailed information on parental caregiving behavior from the Child Supplement of the National Longitudinal Survey of Youth 1979 (NLSY79-CH) is used to examine how the quality of a child's home environment varies with the unemployment rate.

This paper is related to a substantial literature studying how macroeconomic fluctuations affect health outcomes.[1] Ruhm (2000) uncovers a procyclical relationship between mortality and unemployment, although suicides rise during recessions. Using data on babies born in the late twentieth century, Dehejia and Lleras-Muney (2004) find that infant health tends to improve when state unemployment rates increase. Based on a sample of individuals born in the Netherlands between 1812 and 1912, van den Berg et al. (2006) observe that children born during recessions display higher mortality later in life. The current paper extends this line of research in at least two directions. First, whereas much existing work focuses on health outcomes, the current paper analyzes economic variables such as schooling, income, and employment. Second, whereas some existing work has documented the impact of recessions on health behaviors, the current paper examines how parental caregiving and home environments change over the business cycle.

The investigation in this paper begins with an analysis using Census data of the relationship between childhood economic conditions and labor market performance in adulthood. The average unemployment rate between the year before one's birth and the year of one's fifteenth birthday is used to appraise the state of the macroeconomy during one's childhood.[2] Variation in both the national and the state unemployment rate is examined. Specifications using the national unemployment rate as a regressor control for basic demographic variables as well as state fixed effects and current economic conditions. Regressions involving the state unemployment rate also account for national cohort effects and linear state trends. In addition, several robustness checks are performed so as to determine the sensitivity of the results to changes in the measurement of economic fluctuations and the construction of the estimation sample.

The next segment of this paper uses data from the PSID to study whether differences over the business cycle in the quality of parents raising children are likely to explain the observed impacts of childhood conditions on adult outcomes.[3] A number of exercises are conducted for this

[1] Other relevant studies include: Beaudry and DiNardo (1991), who analyze the impact of the lowest unemployment rate since beginning a job on the wage; Malmendier and Nagel (2011), who examine the influence of stock market returns during one's adult lifetime on risk preferences; and Oreopoulos et al. (2012), who investigate the long-term effects of an economic downturn at college graduation on earnings.

[2] Some authors have argued that early childhood is an especially crucial period for human capital formation (Almond and Currie, 2011; Cunha and Heckman, 2007). Therefore, estimates for the impacts of unemployment rates at different stages of development are also presented.

[3] See Dehejia and Lleras-Muney (2004) for a theoretical and empirical discussion of how the unemployment rate affects the characteristics of women selecting to give birth.

purpose. First, I examine how the estimated coefficient on the unemployment rate changes after controlling for parental background variables such as mother's and father's year of birth, education, and occupation. Second, I document the relationship between the unemployment rate during one's childhood and the schooling, employment, and income of one's parents. Third, I estimate family fixed-effects models using sibling data so as to account for the influence of parental background on the results.

The final component of this paper analyzes information from the NLSY79-CH on home environments and caregiving practices in order to illustrate a possible mechanism through which childhood economic conditions can affect the stock of human capital in adulthood. A series of results are presented. I start by estimating the impact of the current unemployment rate on an aggregate measure for the quality of a child's home environment. Next, I investigate whether the estimates are sensitive to the inclusion of maternal background variables, and I characterize the relationship between the unemployment rate and the attributes of mothers raising children. To disentangle the causal effect of economic conditions from changes in parental quality, I compute family and person fixed-effects estimates for the impact of the unemployment rate on a child's home environment. Finally, I examine how specific parenting behaviors vary over the course of the business cycle, and I discuss how economic conditions around the time of birth affect prenatal and postnatal care.

The remainder of this paper is organized as follows. Section 2.2 summarizes the information on unemployment rates as well as the data from the Census, PSID, and NLSY79-CH. Section 2.3 discusses the basic estimates from each dataset and presents various supplemental results and robustness checks. Section 2.4 contains some concluding remarks.

2.2 Data

This section outlines the construction of the datasets in the paper. Section 2.2.1 describes the source of the national and state unemployment rates used in the empirical analysis. Sections 2.2.2, 2.2.3, and 2.2.4 document the main estimation samples for the Census, PSID, and NLSY79-CH, respectively.[4]

[4]Further information about each sample used in the analysis is located in the notes to the tables.

2.2.1 Unemployment Rate Series

A national unemployment rate series from 1890 to 2010 is compiled as follows. For each year from 1941 to 2010, the annual average unemployment rate is obtained from the Bureau of Labor Statistics (BLS).[5] Between 1931 and 1940, the estimates of the unemployment rate from Coen (1973) are used. Between 1890 and 1930, the unemployment rate series constructed by Romer (1986) is employed.

A state unemployment rate series from 1947 to 2009 is generated as follows.[6] For each year from 1976 to 2009, the annual average unemployment rate for each state is obtained from the BLS. Because the BLS does not provide state unemployment rates prior to 1976, yearly information on the rate of insured unemployment is obtained for each state from ET Financial Data Handbook 349. The rate of insured unemployment is available for all states from 1947 to 2009.[7] In order to estimate the unemployment rate for each state between 1947 and 1975, the average annual unemployment rate for a given state is regressed on the rate of insured unemployment and a linear trend in year using the observations on that state between 1976 and 2009.[8] The estimated regression equation for that state is then applied to the rates of insured unemployment to predict the average annual unemployment rates between 1947 and 1975.

In addition, some robustness checks later in the paper replace the unemployment rate with the employment-to-population ratio as a measure of macroeconomic conditions. Because national and state employment-to-population ratios are available from the BLS beginning respectively in 1948 and 1976, the values of these variables in earlier years are estimated as follows. The national employment-to-population ratio is regressed on the national unemployment rate and a linear trend in year using observations from 1948 to 2010, and the estimated regression equation is applied to historical data on the national unemployment rate to predict the national employment-to-population ratios between 1890 and 1947. The employment-to-population ratio for a given state is regressed

---

[5]The national unemployment rate covers individuals 16 years old and above from 1948 to 2010 and individuals 14 years old and above from 1941 to 1947.

[6]The District of Columbia is included as a state.

[7]Only three states—Georgia, Hawaii, and Oregon—have data on the rate of insured unemployment before 1947.

[8]The coefficient on the linear trend in year is significantly positive for five states and significantly negative for eleven states.

on the rate of insured unemployment and a linear trend in year using the observations on that state from 1976 to 2009, and the estimated regression equation is applied to the rates of insured unemployment to predict the employment-to-population ratios for that state between 1947 and 1975.

2.2.2 Census Sample

To document the relationship between unemployment rates in childhood and adult economic outcomes, I construct a sample using data from the Integrated Public Use Microdata Series (IPUMS) for the 1960, 1970, and 1980 Censuses.[9] The dataset is restricted to individuals aged between 30 and 65 at the end of the Census year who have data on educational attainment, working last year, employment status, labor force status, and wage income. In addition, only persons born in one of the fifty states or the District of Columbia are included. Consequently, the sample used for the analysis of national unemployment rates in childhood contains data on individuals born between 1895 and 1930 in the 1960 Census, between 1905 and 1940 in the 1970 Census, and between 1915 and 1950 in the 1980 Census. Because state unemployment rates are available for all states only from 1947 onwards, the sample used for the analysis of state unemployment rates between the year before one's birth and the year of one's fifteenth birthday contains data on individuals born between 1948 and 1950 in the 1980 Census.

Table 2.1 displays summary statistics for the main samples from the Census. The samples used with national and state unemployment rates contain 12,374,991 and 917,783 observations, respectively. The mean years of birth for the respective samples are 1929 and 1949. Correspondingly, the mean ages are 46 and 31. For the former sample, the average national unemployment rate between the year before one's birth and the year of one's fifteenth birthday has mean 8.14 and standard deviation 3.05. For the latter sample, the average state unemployment rate in this interval has mean 8.80 and standard deviation 1.87. The outcomes used in the analysis are: indicators for high school completion, college graduation, and receipt of some graduate education; indicators for having worked in the past calendar year, currently being in the labor force, and being employed

[9] In particular, I combine: the 1960 1% sample; forms 1 and 2 of the 1970 1% metro, state, and neighborhood samples; and the 1980 5%, 1%, 1% urban/rural, 1% labor market area, and 1% detailed metro/non-metro samples. Information from more recent Censuses is not used because precise information on year of birth is not available.

at present; and indicators for having both worked in the past calendar year and received a wage income of at least $10,000, $20,000, and $30,000 during that period.[10] Note that the analysis of income levels utilizes joint work-wage outcomes instead of log wages so as to account for selection into employment.[11]

2.2.3 PSID Sample

In order to examine whether the observed relationship between unemployment rates in childhood and adult economic outcomes can be attributed to changes in the characteristics of parents raising children, I begin by constructing a sample of respondents from the 1968 to 2009 waves of the PSID.[12] These individuals are later matched to information on their parents, provided that their parents are also respondents in the PSID. The dataset contains sample family members from both the Survey Research Center (SRC) and Survey of Economic Opportunity (SEO) components of the PSID. The analysis is restricted to individuals with valid data on year of birth who grew up in one of the 50 states or the District of Columbia.

One observation is generated on an individual for each survey year in which the individual is a head or wife between the ages of 30 and 65 as of the end of the year and has data on years of schooling, total hours worked in the past calendar year, total labor income in the past calendar year, and current employment status. Overall, the sample used for the analysis of national unemployment rates in childhood includes observations on individuals with birth years ranging from 1903 to 1979. When analyzing state unemployment rates between the year before one's birth and the year of one's fifteenth birthday, the sample is limited to individuals born between 1948 and 1979, because state unemployment rates are available for all states only from 1947 onwards.

Descriptive statistics for the main samples from the PSID are presented in Table 2.2. The samples used with national and state unemployment rates comprise 150,604 observations on 11,802 individuals and 67,122 observations on 7,138 individuals, respectively. The mean years of birth

[10] The income figures are expressed in 1982-1984 terms.

[11] Other methods of accommodating the employment decision include the use of a median regression or a selection correction. However, such procedures are difficult to justify here because they usually rely on an assumption about the wage offers of nonparticipants relative to participants or the existence of a variable affecting participation but not wage offers.

[12] The data from the PSID are annual from 1968 to 1997 and biennial thereafter.

Table 2.1: Descriptive Statistics for Observations from Census Sample

	Born Between 1895 and 1950	Born Between 1948 and 1950
Basic Demographics		
Pct. Black	10.15	12.04
Pct. Hispanic	2.44	3.76
Pct. White	88.78	86.42
Pct. Female	51.60	50.60
Region of Residence		
Pct. Northcentral	31.84	30.58
Pct. Northeast	25.19	23.13
Pct. South	33.39	32.42
Pct. West	9.57	13.87
Mean (S.D.) Year Born	1929.48 (12.08)	1949.02 (0.82)
Mean (S.D.) Age	46.06 (10.41)	30.99 (0.82)
Unemployment Rate		
National U.E. Rate		
Mean (S.D.) at Age -1	8.04 (5.57)	——
Mean (S.D.) at Age 0	8.10 (5.58)	——
Mean (S.D.) btw. Ages 1 and 5	8.04 (5.09)	——
Mean (S.D.) btw. Ages 6 and 10	8.16 (5.34)	——
Mean (S.D.) btw. Ages 11 and 15	8.23 (5.32)	——
State U.E. Rate		
Mean (S.D.) at Age -1	——	8.67 (3.60)
Mean (S.D.) at Age 0	——	9.38 (3.43)
Mean (S.D.) btw. Ages 1 and 5	——	8.01 (2.04)
Mean (S.D.) btw. Ages 6 and 10	——	9.13 (2.23)
Mean (S.D.) btw. Ages 11 and 15	——	9.16 (1.83)
Schooling		
Pct. High School and Above	65.25	86.24
Pct. College and Above	14.78	24.58
Pct. Some Graduate School	6.98	11.00
Employment		
Pct. Worked Last Year	73.84	82.28
Pct. in Labor Force	70.43	78.85
Pct. Currently Employed	67.63	74.36
Wage Income		
Pct. Worked and Income $\geq$ \$10K	49.04	55.70
Pct. Worked and Income $\geq$ \$20K	28.32	28.25
Pct. Worked and Income $\geq$ \$30K	12.59	9.62
Sample Size		
Observations	12,374,991	917,783

Note: The summary statistics above are based on the main estimation sample for the Census. Wage income is deflated using the CPI with 1982-1984 as the base period. National and state unemployment rates are constructed as described in the text.

for the respective samples are 1944 and 1957. Correspondingly, the mean ages are 45 and 38. For the former sample, the average national unemployment rate between the year before one's birth and the year of one's fifteenth birthday has mean 6.85 and standard deviation 2.87. For the latter sample, the average state unemployment rate in this interval has mean 7.94 and standard deviation 1.70. The outcomes examined are: indicators for high school completion, college graduation, and receipt of some graduate training; indicators for having worked in the past calendar year, currently being in the labor force, and being employed at present; and indicators for having both worked in the past calendar year and received at least $10,000, $20,000, and $30,000 in labor income during that period.[13]

2.2.4 NLSY79-CH Sample

In order to understand how parental caregiving behavior changes with the unemployment rate, I construct a sample of individuals from the 1986 to 2008 waves of the NLSY79-CH, which surveys children born to female participants in the NLSY79.[14] The restricted-access geocode files for the NLSY79 and NLSY79-CH are obtained so as to match respondents to state-level data on the unemployment rate.[15]

Information from the Home Observation for Measurement of the Environment-Short Form (HOME-SF) inventory is used to assess the quality of each child's household surroundings.[16] The scores on the HOME-SF inventory are based on both parental reports and interviewer observations. The topics covered by the HOME-SF vary with each child's developmental level: infant/toddler (part A, ages 0-2), early childhood (part B, ages 3-5), middle childhood (part C, ages 6-9), and early adolescence (part D, ages 10-14). Examples of items on the HOME-SF include: number of children's books and toys at home; frequency of visits to the grocery, theater, and museum; whether the child eats meals with his/her mother and father; whether the child's mother spoke to, caressed,

[13]The income figures are expressed in 1982-1984 terms.

[14]Individuals in the NLSY79-CH are interviewed biennially.

[15]This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

[16]The HOME-SF is a condensed version of the longer Home Observation for Measurement of the Environment (HOME) inventory. The HOME-SF inventory was developed for use in the NLSY79-CH and is also administered in the PSID. See Caldwell and Bradley (2003) for more details on the HOME inventory.

Table 2.2: Descriptive Statistics for Observations from PSID Sample

	Born Between 1903 and 1979	Born Between 1948 and 1979
Basic Demographics		
Pct. Black	34.54	38.92
Pct. Hispanic	2.54	3.08
Pct. White	61.49	56.99
Pct. Female	52.62	55.43
Region of Residence		
Pct. Northcentral	26.43	27.89
Pct. Northeast	17.20	15.72
Pct. South	45.51	43.04
Pct. West	10.87	13.34
Mean (S.D.) Year Born	1943.72 (14.61)	1956.92 (6.91)
Mean (S.D.) Age	44.92 (10.16)	38.42 (6.88)
Unemployment Rate		
National U.E. Rate		
Mean (S.D.) at Age -1	7.02 (5.12)	——
Mean (S.D.) at Age 0	7.05 (5.14)	——
Mean (S.D.) btw. Ages 1 and 5	7.11 (4.67)	——
Mean (S.D.) btw. Ages 6 and 10	6.93 (4.46)	——
Mean (S.D.) btw. Ages 11 and 15	6.42 (3.84)	——
State U.E. Rate		
Mean (S.D.) at Age -1	——	8.34 (2.93)
Mean (S.D.) at Age 0	——	8.34 (2.92)
Mean (S.D.) btw. Ages 1 and 5	——	8.28 (2.28)
Mean (S.D.) btw. Ages 6 and 10	——	8.08 (2.29)
Mean (S.D.) btw. Ages 11 and 15	——	7.29 (1.88)
Schooling		
Pct. High School and Above	75.86	89.79
Pct. College and Above	19.13	23.07
Pct. Some Graduate School	7.91	9.35
Employment		
Pct. Worked Last Year	82.17	87.15
Pct. in Labor Force	78.65	85.30
Pct. Currently Employed	74.25	78.83
Wage Income		
Pct. Worked and Income $\geq$ \$10K	56.70	60.59
Pct. Worked and Income $\geq$ \$20K	30.92	31.61
Pct. Worked and Income $\geq$ \$30K	14.44	13.95
Sample Size		
Individuals	11,802	7,138
Observations	150,604	67,122

Note: The summary statistics above are based on the main estimation sample for the PSID. Wage income is deflated using the CPI with 1982-1984 as the base period. National and state unemployment rates are constructed as described in the text.

or spanked the child during the interview; how often the child spends time with his/her father; whether the child's mother helps teach the child numbers, letter, colors, and shapes; whether the child is expected to make his/her bed, clean up after him/herself, and perform regular housekeeping tasks; whether the child's home appears to be well lighted, clean, and free of trash. The HOME inventory has been widely employed in the child psychology literature to study how the family setting affects cognitive and behavioral development.[17]

The sample from the NLSY79-CH contains individuals whose mother belongs to the cross-sectional or supplemental sample of the NLSY79. The dataset is restricted to observations on children who live in one of the fifty states or the District of Columbia and are aged between 0 and 15 as of the end of the survey year. Each observation is classified into one of four categories, depending on which age-appropriate part of the HOME-SF inventory was administered to the child in that survey year. Each category includes only observations in which the child has valid data on the total, cognitive stimulation, and emotional support raw scores for the applicable part of the HOME-SF inventory. In addition, many of the items used to compute the scores are individually analyzed in order to further investigate how parenting behavior changes over the business cycle. Therefore, observations on individuals with missing data on these items are excluded from the analysis.[18]

Table 2.3 summarizes the main sample of children from the NLSY79-CH. The datasets for parts A, B, C, and D of the HOME-SF inventory respectively contain 6,505 observations on 5,280 individuals, 8,187 observations on 6,386 individuals, 11,749 observations on 7,523 individuals, and 11,259 observations on 6,620 individuals. The mean survey years for the respective samples are 1991, 1993, 1995, and 1998. Correspondingly, the mean ages are 1.6, 4.5, 8.0, and 12.2. For the respective parts, the current national unemployment rate has means 6.01, 5.88, 5.76, 5.52 and standard deviations 0.95, 0.95, 0.99, and 0.96, and the current state unemployment rate has means

[17]For example, see Elardo et al. (1977), Bradley and Caldwell (1980), and Bradley et al. (1988).

[18]The specific items examined are as follows with the relevant parts given in parentheses: number of children's books (A, B, C, D); frequency of being read to (A, B, C); frequency of grocery visits (A); number of cuddly or role-playing toys (A); number of push or pull toys (A); frequency of seeing father (A, B); frequency of eating with parents (A, B, C, D); frequency of being spanked (A, B, C, D); number of magazines (B); presence of tape recorder (B); frequency of museum visits (B, C, D); hours of television (B); presence of musical instrument (C, D); receipt of daily newspaper (C, D); frequency of theater visits (C, D); enrollment in special lessons (C, D); frequency of interacting with father (C, D); discussion of television shows (C, D).

6.10, 5.93, 5.82, and 5.56 and standard deviations 1.64, 1.59, 1.57, and 1.42. The main outcome variables are the total, cognitive stimulation, and emotional support scores on each part of the HOME-SF inventory.[19]

2.3 Results

This section discusses the results from the analysis of unemployment rates in childhood. Sections 2.3.1, 2.3.2, and 2.3.3 present the findings from the Census, PSID, and NLSY79-CH, respectively.

2.3.1 Census Results

Section 2.3.1 describes the basic estimates for the Census sample. Section 2.3.1 reports some robustness checks for the Census results.

Basic Estimates

The upper panel of Table 2.4 displays estimates for the impact of national unemployment rates between the year before one's birth and the year of one's fifteenth birthday on schooling, employment, and income between the ages of 30 and 65. The specifications control for race, gender, state of birth, age dummies, and indicators for survey year.[20] In general, the coefficients on national unemployment rates in childhood are moderate in size and precisely estimated. For schooling outcomes, the average national unemployment rate in childhood has an insignificantly positive impact on high school completion and a significantly negative impact on graduation from college and attendance in graduate school. For employment outcomes, a higher average national unemployment rate in childhood significantly lowers the likelihoods of having worked in the previous year, of currently participating in the labor force, and of being employed at present. For income outcomes, the

[19] Although Table 2.3 reports summary statistics for the raw scores on the HOME-SF inventory, the regression analysis uses standardized scores so as to facilitate interpretation of the results. Furthermore, several items from each part of the HOME-SF inventory are analyzed as discussed in the text but omitted from the tables because of space constraints.

[20] The standard errors are clustered by year of birth, because the national unemployment rate in childhood is the same for all individuals with the same year of birth.

Table 2.3: Descriptive Statistics for Observations from NLSY79-CH Sample

	HOME-SF Inventory			
	Part A: Infant/Toddler	Part B: Early Childhood	Part C: Middle Childhood	Part D: Early Adolescence
Basic Demographics				
Pct. Black	24.77	24.44	27.46	29.19
Pct. Hispanic	18.80	19.08	19.41	20.46
Pct. White	56.43	56.48	53.13	50.35
Pct. Female	49.19	49.77	49.28	49.86
Region of Residence				
Pct. Northcentral	27.70	27.48	26.84	27.17
Pct. Northeast	16.89	16.17	15.46	14.38
Pct. South	34.80	35.98	37.56	38.32
Pct. West	20.60	20.36	20.14	20.13
Mean (S.D.) Age	1.58 (0.93)	4.51 (0.94)	8.01 (1.29)	12.15 (1.36)
Mean (S.D.) Year	1991.34 (4.67)	1992.78 (5.22)	1994.62 (5.84)	1997.62 (5.47)
Unemployment Rate				
Mean (S.D.) National U.E. Rate	6.01 (0.95)	5.88 (0.95)	5.76 (0.99)	5.52 (0.96)
Mean (S.D.) State U.E. Rate	6.10 (1.64)	5.93 (1.59)	5.82 (1.57)	5.56 (1.42)
HOME-SF Inventory				
Mean (S.D.) Total Raw Score	141.53 (23.98)	206.37 (35.47)	201.17 (36.67)	204.48 (34.92)
Mean (S.D.) Cognitive Stimulation Raw Score	68.03 (15.63)	117.78 (22.19)	99.15 (24.25)	93.86 (22.83)
Mean (S.D.) Emotional Support Raw Score	73.53 (14.52)	88.54 (19.79)	102.02 (19.88)	110.61 (19.85)
Sample Size				
Individuals	5,280	6,386	7,523	6,620
Observations	6,505	8,187	11,749	11,259

Note: The summary statistics above are based on the main estimation sample for the NLSY79-CH. Parts A, B, C, and D of the HOME-SF inventory are generally administered to children aged 0-2, 3-5, 6-9, and 10-14 years, respectively. National and state unemployment rates are constructed as described in the text.

impact of a higher national unemployment rate in childhood is more complex, lowering the probability of having worked and earned at least \$10,000 and increasing the probabilities of having worked and earned at least \$20,000 and \$30,000.

The lower panel of Table 2.4 replicates the preceding analysis using state instead of national unemployment rates in childhood. Recall that state unemployment rates are available for only a small subset of the sample used with national unemployment rates. The estimates control for race, gender, birth state, age dummies, and a linear trend in age specific to each birth state.[21] Overall, the coefficients on state unemployment rates in childhood are large in magnitude but imprecisely estimated. Significant negative impacts are found on the probabilities of currently participating in the labor force, being employed at present, and having worked and earned at least \$10,000. Most of the point estimates are negative in sign, and no significant positive effects are obtained.

In addition, Table 2.4 provides estimates for the impacts of average national and state unemployment rates at different stages of child development: prenatal (age -1), infancy (age 0), early childhood (ages 1-5), middle childhood (ages 6-10), and early adolescence (ages 11-15). Both significantly positive and negative impacts are found for the sample used with national unemployment rates. For the sample used with state unemployment rates, the estimated impacts are typically negative. The results provide little evidence suggesting that high unemployment rates earlier in childhood have a larger negative impact on adult economic outcomes than high unemployment rates later in childhood.

As a preliminary test of whether changes in the characteristics of parents raising children can explain the observed impacts of state unemployment rates in childhood, I use data from the 1960 Census to construct a sample of parents between the ages of 30 and 65 with children born between 1948 and 1950.[22] I then regress indicators for mother's and father's schooling, employment, and income on the average state unemployment rate between the year before the child's birth and the year

[21]The standard errors are clustered by state of birth, in order to account for serial correlation across birth years among individuals born in the same state. See Bertrand et al. (2004) for a discussion of how serial correlation affects the standard errors for differences-in-differences estimates.

[22]The results described here are not included in the tables but are available from the author on request. Because of data limitations, a similar test is not performed using national instead of state unemployment rates in childhood. In particular, precise information on year of birth is typically unavailable in earlier Census years, and individuals do not have data on their parents if they have left the parental home. Consequently, insufficiently many cohorts are available for estimating the relationship between national unemployment rates in childhood and parental characteristics.

Table 2.4: Relationship of Unemployment Rates in Childhood to Schooling, Employment, and Income for Census Sample

	H.S. Diploma	College Degree	Grad. School	Worked Last Yr.	In Labor Force	Currently Employed	Worked & Y ≥ $10K	Worked & Y ≥ $20K	Worked & Y ≥ $30K
					Born Between 1895 and 1950				
					Average National Unemployment Rate in Childhood				
U.E. Rate btw. Ages -1 and 15	0.0005 (0.0005)	-0.0026 (0.0003)	-0.0021 (0.0002)	-0.0076 (0.0004)	-0.0079 (0.0004)	-0.0072 (0.0004)	-0.0029 (0.0002)	0.0002 (0.0002)	0.0012 (0.0002)
					Average National Unemployment Rate at Different Stages of Childhood				
U.E. Rate at Age -1	-0.0004 (0.0003)	-0.0019 (0.0005)	-0.0009 (0.0002)	-0.0019 (0.0008)	-0.0019 (0.0009)	-0.0017 (0.0008)	-0.0007 (0.0004)	0.0011 (0.0004)	0.0016 (0.0004)
U.E. Rate at Age 0	0.0003 (0.0003)	0.0005 (0.0006)	0.0003 (0.0002)	-0.0002 (0.0010)	-0.0001 (0.0010)	0.0000 (0.0009)	0.0000 (0.0005)	0.0002 (0.0006)	-0.0003 (0.0005)
U.E. Rate btw. Ages 1 and 5	0.0012 (0.0002)	-0.0012 (0.0003)	-0.0010 (0.0001)	-0.0026 (0.0006)	-0.0028 (0.0007)	-0.0026 (0.0006)	-0.0012 (0.0003)	-0.0009 (0.0003)	-0.0003 (0.0003)
U.E. Rate btw. Ages 6 and 10	-0.0014 (0.0003)	0.0000 (0.0002)	-0.0001 (0.0001)	-0.0012 (0.0004)	-0.0011 (0.0004)	-0.0009 (0.0004)	-0.0003 (0.0002)	0.0010 (0.0002)	0.0011 (0.0002)
U.E. Rate btw. Ages 11 and 15	0.0018 (0.0002)	-0.0019 (0.0002)	-0.0014 (0.0001)	-0.0040 (0.0003)	-0.0043 (0.0004)	-0.0040 (0.0003)	-0.0015 (0.0002)	-0.0006 (0.0002)	-0.0003 (0.0001)
Birth Years					———56———				
Observations					———12,374,991———				
					Born Between 1948 and 1950				
					Average State Unemployment Rate in Childhood				
U.E. Rate btw. Ages -1 and 15	0.0110 (0.0302)	-0.0051 (0.0385)	-0.0336 (0.0204)	-0.0643 (0.0482)	-0.0941 (0.0424)	-0.1001 (0.0437)	-0.0797 (0.0372)	-0.0447 (0.0427)	-0.0060 (0.0275)
					Average State Unemployment Rate at Different Stages of Childhood				
U.E. Rate at Age -1	0.0022 (0.0020)	-0.0007 (0.0027)	-0.0022 (0.0017)	-0.0030 (0.0035)	-0.0049 (0.0031)	-0.0063 (0.0031)	-0.0061 (0.0026)	-0.0006 (0.0023)	0.0021 (0.0021)
U.E. Rate at Age 0	0.0006 (0.0020)	-0.0010 (0.0024)	-0.0018 (0.0014)	-0.0028 (0.0025)	-0.0049 (0.0030)	-0.0055 (0.0026)	-0.0050 (0.0023)	-0.0036 (0.0018)	0.0001 (0.0016)
U.E. Rate btw. Ages 1 and 5	-0.0017 (0.0118)	-0.0029 (0.0119)	-0.0029 (0.0092)	-0.0072 (0.0129)	-0.0141 (0.0113)	-0.0161 (0.0128)	-0.0191 (0.0133)	-0.0169 (0.0111)	0.0055 (0.0067)
U.E. Rate btw. Ages 6 and 10	-0.0075 (0.0127)	-0.0133 (0.0127)	-0.0090 (0.0103)	-0.0098 (0.0144)	-0.0234 (0.0125)	-0.0231 (0.0138)	-0.0317 (0.0138)	-0.0386 (0.0116)	0.0012 (0.0077)
U.E. Rate btw. Ages 11 and 15	-0.0044 (0.0098)	-0.0070 (0.0109)	-0.0125 (0.0110)	-0.0192 (0.0125)	-0.0257 (0.0118)	-0.0254 (0.0155)	-0.0396 (0.0120)	-0.0193 (0.0165)	0.0121 (0.0083)
Birth States					———51———				
Observations					———917,783———				

Note: The main estimation sample for the Census is used to generate the estimates above. The specifications in the upper panel contain indicator variables for race, gender, birth state, survey year, and age at the end of the survey year. The specifications in the lower panel control for race, gender, birth state, age dummies, and a linear trend in age specific to each birth state. Huber-White standard errors, clustered by birth year in the upper panel and by birth state in the lower panel, are reported in parentheses.

of the child's fifteenth birthday as well as controls for parent's race, child's state of birth, dummy variables for parent's age, fixed effects for child's birth year, and a linear trend in child's year of birth specific to each state of birth. In general, I do not find a significant relationship between state unemployment rates and parental characteristics. Nonetheless, the estimates for the coefficients on state unemployment rates are often imprecise. Hence, it is unclear whether changes over the business cycle in the characteristics of parents raising children may be generating the observed relationship between state unemployment rates in childhood and adult economic outcomes.[23]

Robustness Checks

A few robustness checks are conducted to assess the sensitivity of the results to changes in the sample used for estimation and the measurement of economic conditions.[24] First, I replicate the results in the bottom panel of Table 2.4 using the raw data on the rate of covered unemployment for each state instead of the estimated values of the state unemployment rate to generate the regressors.[25] This substitution does not qualitatively change the results. Second, I perform the regressions in Table 2.4 using the employment-to-population ratio instead of the unemployment rate as an indicator of economic conditions.[26] This modification does not substantially affect the conclusions of the analysis. Third, I estimate the specifications in the upper panel of Table 2.4 using only the 1% samples from the 1960, 1970, and 1980 Censuses.[27] This restriction does not drastically alter the basic pattern of estimates. However, the national unemployment rate between the year before one's birth and the year of one's fifteenth birthday now has a significantly positive impact on high school completion as well as the probability of having worked and earned at least

[23]Section 2.3.2 provides a more complete analysis of the relationship between childhood unemployment rates and parental characteristics based on data from the PSID.

[24]These results are omitted from the tables but are available from the author on request.

[25]Recall from section 2.2.1 that the rate of covered unemployment from ET Financial Data Handbook 349 is used to estimate the annual state unemployment rates between 1947 and 1975, because the BLS does not provide annual state unemployment rates prior to 1976.

[26]As noted by Dehejia and Lleras-Muney (2004), the use of the employment-to-population ratio instead of the unemployment rate avoids measurement error in determining the size of the labor force and the number of unemployed workers.

[27]In the original sample from the upper panel of Table 2.4, individuals in 1980 are overrepresented relative to individuals in 1960 and 1970, and individuals in 1970 are overrepresented relative to individuals in 1960.

$20,000.

2.3.2 PSID Results

The basic estimates for the PSID sample are presented in section 2.3.2, and some robustness checks for the PSID results are summarized in section 2.3.2.

Basic Estimates

Table 2.5 reproduces the analysis in Table 2.4 using the PSID instead of the Census sample.[28] The upper panel displays the impacts of national unemployment rates between the year before one's birth and the year of one's fifteenth birthday on schooling, employment, and income between the ages of 30 and 65. The estimates control for race, gender, childhood state, age dummies, and indicators for survey year.[29] The point estimate for the coefficient on the average national unemployment rate in childhood is negative for each outcome examined. Except for the probability of having worked and earned at least $30,000, all of the deviations from zero are statistically significant.

The impacts of state unemployment rates in childhood on adult economic outcomes are exhibited in the lower panel. The estimates control for race, gender, childhood state, age dummies, fixed effects for survey year, indicator variables for year born, and a linear trend in year born specific to each childhood state.[30] The results are mixed with the point estimates being positive for some outcomes and negative for other outcomes. None of the observed deviations from zero are statistically significant, except for the positive impact on the probability of having worked and earned at least $30,000.

In addition, Table 2.5 presents estimates in which a separate coefficient is computed for the average national or state unemployment rate at each stage of childhood. On the whole, there is little evidence indicating that high unemployment rates earlier in childhood are more detrimental than high unemployment rates later in childhood.

---

[28] Although the sample sizes are smaller for the PSID than for the Census, the PSID sample covers a larger number of birth cohorts than the Census sample.

[29] The standard errors are clustered by year of birth.

[30] The standard errors are clustered by childhood state.

Table 2.5: Relationship of Unemployment Rates in Childhood to Schooling, Employment, and Income for PSID Sample

	H.S. Diploma	College Degree	Grad. School	Worked Last Yr.	In Labor Force	Currently Employed	Worked & Y ≥ \$10K	Worked & Y ≥ \$20K	Worked & Y ≥ \$30K
					Born Between 1903 and 1979				
					Average National Unemployment Rate in Childhood				
U.E. Rate btw.	-0.0092	-0.0078	-0.0037	-0.0073	-0.0075	-0.0078	-0.0064	-0.0037	-0.0014
Ages -1 and 15	(0.0025)	(0.0019)	(0.0013)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0019)	(0.0014)
					Average National Unemployment Rate at Different Stages of Childhood				
U.E. Rate at	-0.0054	-0.0069	-0.0025	-0.0036	-0.0032	-0.0036	-0.0036	-0.0016	0.0005
Age -1	(0.0032)	(0.0021)	(0.0014)	(0.0017)	(0.0019)	(0.0021)	(0.0028)	(0.0024)	(0.0020)
U.E. Rate at	0.0009	0.0031	0.0013	0.0016	0.0017	0.0024	0.0007	-0.0001	-0.0013
Age 0	(0.0031)	(0.0022)	(0.0014)	(0.0019)	(0.0021)	(0.0022)	(0.0030)	(0.0026)	(0.0024)
U.E. Rate btw.	-0.0012	-0.0027	-0.0014	-0.0025	-0.0033	-0.0035	-0.0014	-0.0022	-0.0003
Ages 1 and 5	(0.0028)	(0.0012)	(0.0009)	(0.0015)	(0.0015)	(0.0016)	(0.0023)	(0.0018)	(0.0017)
U.E. Rate btw.	-0.0025	-0.0016	-0.0007	-0.0017	-0.0012	-0.0014	-0.0020	0.0005	-0.0001
Ages 6 and 10	(0.0023)	(0.0011)	(0.0008)	(0.0012)	(0.0012)	(0.0012)	(0.0017)	(0.0015)	(0.0014)
U.E. Rate btw.	-0.0064	-0.0033	-0.0017	-0.0033	-0.0032	-0.0033	-0.0019	-0.0018	-0.0002
Ages 11 and 15	(0.0023)	(0.0011)	(0.0008)	(0.0011)	(0.0012)	(0.0012)	(0.0013)	(0.0013)	(0.0010)
Birth Years					77				
Individuals					11,802				
Observations					150,604				
					Born Between 1948 and 1979				
					Average State Unemployment Rate in Childhood				
U.E. Rate btw.	-0.0026	-0.0086	-0.0163	-0.0043	-0.0062	0.0040	0.0119	0.0184	0.0138
Ages -1 and 15	(0.0133)	(0.0123)	(0.0100)	(0.0055)	(0.0060)	(0.0071)	(0.0123)	(0.0107)	(0.0070)
					Average State Unemployment Rate at Different Stages of Childhood				
U.E. Rate at	-0.0030	0.0079	0.0017	0.0020	0.0016	0.0032	0.0043	0.0081	0.0090
Age -1	(0.0038)	(0.0050)	(0.0034)	(0.0029)	(0.0032)	(0.0033)	(0.0048)	(0.0051)	(0.0041)
U.E. Rate at	0.0041	-0.0035	0.0032	0.0014	0.0037	0.0037	0.0023	0.0000	0.0014
Age 0	(0.0030)	(0.0052)	(0.0040)	(0.0030)	(0.0035)	(0.0037)	(0.0043)	(0.0046)	(0.0023)
U.E. Rate btw.	0.0028	-0.0086	-0.0174	-0.0031	-0.0058	-0.0020	0.0021	-0.0008	-0.0049
Ages 1 and 5	(0.0062)	(0.0094)	(0.0052)	(0.0037)	(0.0037)	(0.0042)	(0.0074)	(0.0076)	(0.0049)
U.E. Rate btw.	-0.0092	0.0075	0.0050	0.0023	0.0012	0.0067	0.0067	0.0140	0.0134
Ages 6 and 10	(0.0050)	(0.0088)	(0.0048)	(0.0029)	(0.0032)	(0.0038)	(0.0073)	(0.0062)	(0.0043)
U.E. Rate btw.	0.0025	-0.0046	-0.0016	-0.0044	-0.0025	-0.0038	0.0004	0.0043	0.0051
Ages 11 and 15	(0.0047)	(0.0100)	(0.0069)	(0.0029)	(0.0047)	(0.0041)	(0.0062)	(0.0065)	(0.0051)
Childhood States					49				
Individuals					7,138				
Observations					67,122				

Note: The main estimation sample for the PSID is used to generate the estimates above. The specifications in the upper panel contain indicator variables for race, gender, childhood state, survey year, and age at the end of the survey year. The specifications in the lower panel control for race, gender, childhood state, age dummies, fixed effects for survey year, indicator variables for year born, and a linear trend in year born specific to each childhood state. Huber-White standard errors, clustered by birth year in the upper panel and by childhood state in the lower panel, are reported in parentheses.

I now conduct several exercises to determine whether changes over the business cycle in the characteristics of parents raising children are likely to explain the observed relationship between unemployment rates in childhood and adult labor market outcomes. I begin by identifying a subset of individuals from the sample in Table 2.5 whose mother or father has information on important background variables and economic outcomes, and I replicate the analysis from the upper row in each panel of Table 2.5 excluding and including controls for parental background variables such as year born, education, and occupation.[31]

Table 2.6 reports the results of this procedure. Using the average national unemployment rate in childhood as a regressor in the upper panel, the estimated coefficients on the unemployment rate decrease for six outcomes and increase for three outcomes when parental background variables are added to the specification. Both before and after controlling for parental background, the only statistically significant result obtained for this sample is a negative impact on high school completion. Using the average state unemployment rate in childhood as a regressor in the lower panel, the estimated coefficients on the unemployment rate decrease for all nine outcomes as a result of controlling for parental background. Although no statistically significant results are detected when parental background variables are excluded, significant negative impacts are seen on graduation from college and receipt of graduate training after the inclusion of parental background controls. In sum, it appears unlikely that changes over the business cycle in the characteristics of parents raising children are contributing to a negative relationship between childhood unemployment rates and adult economic outcomes.

I next document how the characteristics of parents raising children vary with the unemployment rate. To perform this exercise, I construct a sample consisting of observations between the ages of 30 and 65 on the schooling, employment, and income of mothers and fathers of children in the dataset from Table 2.6.[32] I then regress indicators for each parent's outcomes on the average national or state unemployment rate between the year before the child's birth and the year of the child's fifteenth birthday as well as other control variables.[33] These regressions are presented in

[31]See the note to Table 2.6 for more details on the construction of the sample and the specification being estimated. Observe that the dataset used in Table 2.6 consists of PSID respondents whose mother or father is also a sample member in the PSID.

[32]See the note to Table 2.7 for additional information on the sample selection criteria.

[33]These regressions are run separately for mothers and for fathers. In specifications with the average national

Table 2.6: Relationship of Unemployment Rates in Childhood to Schooling, Employment, and Income for PSID Sample Before and After Controlling for Parental Background Variables

	H.S. Diploma	College Degree	Grad. School	Worked Last Yr.	In Labor Force	Currently Employed	Worked & Y ≥ \$10K	Worked & Y ≥ \$20K	Worked & Y ≥ \$30K
				Has Parent with Data in Survey and Born Between 1903 and 1979					
				Average National Unemployment Rate in Childhood					
U.E. Rate btw. Ages -1 and 15	-0.0265 (0.0101)	-0.0026 (0.0117)	0.0082 (0.0088)	-0.0027 (0.0091)	0.0008 (0.0090)	-0.0014 (0.0109)	-0.0050 (0.0112)	0.0044 (0.0100)	0.0069 (0.0065)
Parental Background	No	No	No	No	No	No	No	No	No
				Average National Unemployment Rate in Childhood					
U.E. Rate btw. Ages -1 and 15	-0.0298 (0.0122)	-0.0052 (0.0146)	0.0051 (0.0101)	-0.0050 (0.0103)	0.0009 (0.0103)	-0.0028 (0.0116)	-0.0057 (0.0137)	0.0057 (0.0147)	0.0098 (0.0086)
Parental Background	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Years					51				
Individuals					6,742				
Observations					64,798				
				Has Parent with Data in Survey and Born Between 1948 and 1979					
				Average State Unemployment Rate in Childhood					
U.E. Rate btw. Ages -1 and 15	-0.0037 (0.0143)	-0.0106 (0.0130)	-0.0158 (0.0098)	-0.0063 (0.0059)	-0.0061 (0.0065)	0.0016 (0.0077)	0.0101 (0.0133)	0.0124 (0.0122)	0.0094 (0.0074)
Parental Background	No	No	No	No	No	No	No	No	No
				Average State Unemployment Rate in Childhood					
U.E. Rate btw. Ages -1 and 15	-0.0129 (0.0131)	-0.0384 (0.0118)	-0.0291 (0.0079)	-0.0092 (0.0051)	-0.0099 (0.0061)	-0.0028 (0.0070)	0.0009 (0.0114)	-0.0036 (0.0105)	-0.0041 (0.0076)
Parental Background	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Childhood States					49				
Individuals					6,439				
Observations					58,642				

Note: The dataset used here is constructed by restricting the main estimation sample for the PSID to observations on individuals whose mother or father has information on first occupation and birth year as well as years of schooling, total hours worked, total labor income, and employment status for some survey year when aged between 30 and 65. Parental background variables are indicators for mother's and father's first occupation, educational attainment, and birth year. The categories for occupation at first regular job are: professional and technical workers; managers, officials, and proprietors; self-employed businessman; clerical and sales workers; craftsmen and foremen; operatives; laborers and service workers; and armed services and protective workers. The categories for educational attainment are: less than high school graduate, high school diploma, some college, college degree, and some post-graduate training. The specifications in the upper panel contain indicator variables for race, gender, childhood state, survey year, and age at the end of the survey year. The specifications in the lower panel control for race, gender, childhood state, age dummies, fixed effects for survey year, indicator variables for year born, and a linear trend in year born specific to each childhood state. Huber-White standard errors, clustered by birth year in the upper panel and by childhood state in the lower panel, are reported in parentheses.

Table 2.7. In general, the coefficient on the unemployment rate is either statistically insignificant or significantly positive. That is, a higher unemployment rate may be associated with an improvement in the quality of parents raising children. Hence, selection over the business cycle into raising children appears unlikely to generate a negative correlation between unemployment rates in childhood and economic performance as an adult.

As a final exercise to control for a possible association between childhood unemployment rates and parental quality, I use sibling data to calculate family fixed-effects estimates for the impact of national unemployment rates in childhood on schooling, employment, and income in adulthood. Table 2.8 contains the results of the estimation.[34] Although the point estimates for the coefficient on the average national unemployment rate between the year before one's birth and the year of one's fifteenth birthday are negative for all but one outcome analyzed, the only statistically significant result is a negative impact on high school completion. In specifications that compute separate coefficients for the average national unemployment rates at different stages of childhood, unemployment rates earlier in childhood do not generally appear to be a bigger determinant of adult economic outcomes than unemployment rates later in childhood. Overall, the results from sibling data are consistent with a negative causal effect of the unemployment rate in childhood on labor market success as an adult, although the estimates are in this case too imprecise to draw a definitive conclusion.

Robustness Checks

This section performs a few robustness checks analogous to those in section 2.3.1.[35] First, the results in the bottom halves of Tables 2.5, 2.6, and 2.7 are replicated using the raw data on the rate of covered unemployment for each state instead of the actual and estimated values of the state unemployment rate to construct the regressors. The basic pattern of estimates in the lower panel

unemployment rate as a regressor, the other control variables are: parent's race, state where child grew up, parental age dummies, fixed effects for survey year of observation on parent, and a linear trend in child's year of birth. In specifications with the average state unemployment rate as a regressor, the other control variables are: parent's race, state where child grew up, parental age dummies, fixed effects for survey year of observation on parent, indicator variables for child's year of birth, and a linear trend in child's year of birth specific to the state where child grew up.

[34]The sample used here comprises all individuals in the dataset from Table 2.5 who have a sibling also belonging to the dataset. The specification includes a gender dummy as well as fixed effects for age level and survey year.

[35]The estimates summarized here are available from the author on request.

Table 2.7: Relationship of Unemployment Rates in Childhood to Parental Schooling, Employment, and Income for PSID Sample

	H.S. Diploma	College Degree	Grad. School	Worked Last Yr.	In Labor Force	Currently Employed	Worked & Y ≥ $10K	Worked & Y ≥ $20K	Worked & Y ≥ $30K
					Mothers of Individuals Born Between 1903 and 1979				
					Average National Unemployment Rate in Youth's Childhood				
U.E. Rate btw. Ages -1 and 15	0.0288 (0.0140)	0.0261 (0.0097)	0.0151 (0.0044)	0.0123 (0.0130)	0.0131 (0.0121)	0.0063 (0.0117)	0.0215 (0.0105)	0.0199 (0.0056)	0.0101 (0.0028)
Birth Years for Youth					49				
Observations on Mother					106,941				
					Fathers of Individuals Born Between 1903 and 1979				
					Average National Unemployment Rate in Youth's Childhood				
U.E. Rate btw. Ages -1 and 15	0.0624 (0.0172)	0.0226 (0.0202)	-0.0163 (0.0133)	0.0068 (0.0081)	0.0102 (0.0099)	0.0077 (0.0098)	0.0128 (0.0113)	0.0213 (0.0148)	0.0192 (0.0139)
Birth Years for Youth					46				
Observations on Father					91,502				
					Mothers of Individuals Born Between 1948 and 1979				
					Average State Unemployment Rate in Youth's Childhood				
U.E. Rate btw. Ages -1 and 15	-0.0112 (0.0171)	0.0034 (0.0082)	0.0020 (0.0029)	0.0121 (0.0074)	0.0127 (0.0082)	0.0097 (0.0080)	0.0051 (0.0088)	0.0052 (0.0057)	0.0006 (0.0027)
Childhood States for Youth					48				
Observations on Mother					104,100				
					Fathers of Individuals Born Between 1948 and 1979				
					Average State Unemployment Rate in Youth's Childhood				
U.E. Rate btw. Ages -1 and 15	0.0584 (0.0203)	0.0238 (0.0192)	0.0235 (0.0154)	0.0065 (0.0088)	0.0046 (0.0089)	0.0050 (0.0110)	0.0330 (0.0140)	0.0148 (0.0222)	0.0094 (0.0237)
Childhood States for Youth					46				
Observations on Father					89,394				

Note: The dataset used for the top (resp. bottom) row of estimates in each half of the table is constructed as follows. First, the main estimation sample for the PSID is restricted to observations on individuals whose mother or father has information on first occupation and birth year as well as years of schooling, total hours worked, total labor income, and employment status for some survey year when aged between 30 and 65. Second, every individual in the resulting sample is matched to the available information on her mother (resp. father). Third, one observation is generated for each survey year in which her mother (resp. father) has valid data on years of schooling, total hours worked, total labor income, and employment status. The specifications in the upper half of the table contain indicator variables for parent's race, youth's childhood state, parent's survey year, and parent's age at the end of the survey year as well as a linear trend in the youth's birth year. The specifications in the lower half of the table control for parent's race, youth's childhood state, parental age dummies, fixed effects for parent's survey year, indicator variables for youth's birth year, and a linear trend in the youth's birth year specific to the youth's childhood state. Huber-White standard errors, clustered by youth's birth year in the upper half and by youth's childhood state in the lower half, are reported in parentheses.

Table 2.8: Family Fixed-Effects Estimates for Relationship of Unemployment Rates in Childhood to Schooling, Employment, and Income for PSID Sample

	H.S. Diploma	College Degree	Grad. School	Worked Last Yr.	In Labor Force	Currently Employed	Worked & Y $\geq$ \$10K	Worked & Y $\geq$ \$20K	Worked & Y $\geq$ \$30K
					Born Between 1903 and 1979				
					Average National Unemployment Rate in Childhood				
U.E. Rate btw. Ages -1 and 15	-0.0330 (0.0151)	-0.0055 (0.0074)	0.0003 (0.0075)	-0.0121 (0.0121)	-0.0089 (0.0118)	-0.0119 (0.0124)	-0.0209 (0.0172)	-0.0092 (0.0092)	-0.0020 (0.0084)
					Average National Unemployment Rate at Different Stages of Childhood				
U.E. Rate at Age -1	-0.0047 (0.0047)	0.0061 (0.0048)	0.0033 (0.0035)	0.0027 (0.0035)	0.0027 (0.0036)	0.0021 (0.0038)	-0.0029 (0.0045)	-0.0002 (0.0045)	0.0025 (0.0035)
U.E. Rate at Age 0	-0.0013 (0.0049)	-0.0124 (0.0057)	-0.0066 (0.0043)	-0.0023 (0.0039)	-0.0006 (0.0040)	-0.0007 (0.0043)	0.0011 (0.0052)	0.0003 (0.0049)	-0.0051 (0.0041)
U.E. Rate btw. Ages 1 and 5	-0.0035 (0.0076)	0.0027 (0.0057)	0.0006 (0.0054)	-0.0046 (0.0055)	-0.0066 (0.0054)	-0.0066 (0.0056)	-0.0106 (0.0072)	-0.0081 (0.0058)	0.0011 (0.0063)
U.E. Rate btw. Ages 6 and 10	-0.0193 (0.0066)	0.0049 (0.0057)	0.0054 (0.0044)	-0.0085 (0.0060)	-0.0051 (0.0058)	-0.0079 (0.0065)	-0.0021 (0.0088)	0.0040 (0.0054)	0.0020 (0.0044)
U.E. Rate btw. Ages 11 and 15	-0.0071 (0.0073)	0.0071 (0.0067)	0.0103 (0.0058)	-0.0006 (0.0049)	-0.0004 (0.0052)	0.0006 (0.0055)	-0.0014 (0.0071)	0.0001 (0.0061)	0.0037 (0.0055)
Families					1,897				
Individuals					5,703				
Observations					58,849				

Note: The dataset used here is constructed by restricting the main estimation sample for the PSID to observations on individuals with a sibling also in the sample for some year. The specifications include fixed effects for survey year and dummy variables for age at the end of the survey year. Huber-White standard errors, clustered by family, are reported in parentheses.

of Table 2.5 remains intact after this substitution. However, the positive impact of average state unemployment on having worked and earned at least $30,000 is no longer statistically significant. The main conclusions from the bottom halves of Tables 2.6 and 2.7 are also unchanged when the unemployment rate is replaced with the rate of covered unemployment.

Second, the regressions in Tables 2.5, 2.6, 2.7, and 2.8 are conducted using the employment-to-population ratio instead of the unemployment rate as a measure of macroeconomic conditions. The findings from Table 2.5 change little as a result of this modification. In the lower panel, however, the positive impact of worse childhood economic conditions on having worked and earned at least $30,000 loses significance, and the negative impact of worse childhood economic conditions on receipt of graduate training gains significance. Although the main conclusions from the bottom halves of Tables 2.6 and 2.7 are unaffected, some changes in the top halves of these tables should be noted. When the employment-to-population ratio is used in the top halves of Tables 2.6 and 2.7, the negative impact of worse childhood economic conditions on adult outcomes frequently weakens when parental background variables are included in the specification, and worse economic conditions while raising a child are occasionally associated with significantly lower schooling and income among parents.

By contrast, the negative impact of worse childhood economic conditions often becomes stronger in the family fixed-effects estimates from Table 2.8 when the unemployment rate is replaced with the employment-to-population ratio. Even though the impact on high school completion is no longer statistically significant, worse childhood economic conditions are now seen to have a very significant negative impact on college graduation, receipt of graduate training, and having worked and earned at least $20,000. On balance, there continues to be evidence that the negative impact of worse childhood economic conditions on some adult outcomes cannot be fully explained by changes over the business cycle in the background characteristics of parents raising children.

Third, the specifications in Tables 2.5, 2.6, 2.7, and 2.8 are estimated only for members of the nationally representative SRC sample of the PSID.[36] The results in Table 2.5 are mostly robust to this sample restriction. However, the negative impacts of the average national unemployment rate on receipt of graduate training and having worked with earnings of at least $20,000 lose

[36]The original dataset combines individuals in the SRC and SEO samples. Low-income households are overrepresented in the SEO sample.

significance in the upper panel, and the negative impact of the average state unemployment rate on having worked in the previous year gains significance in the lower panel.

The findings from Table 2.6 are more complex after the exclusion of the SEO sample. In the upper panel, the impacts of the national unemployment rate on having worked and earned at least $10,000 and $20,000 now become larger in size and are significantly positive when parental background variables are added as regressors. In the lower panel, the basic pattern of estimates is largely robust to the sample restriction. Likewise, the findings from Table 2.7 are more complex after the exclusion of the SEO sample. In the upper half of the table, the impact of the national unemployment rate is now significantly positive, statistically insignificant, or significantly negative depending on the parental outcome examined. In the lower half of the table, the basic pattern of estimates is largely robust to the sample restriction.

The family fixed-effects estimates from Table 2.8 are essentially unaffected by the removal of the SEO sample. The negative impact of the national unemployment rate on high school completion now becomes larger in size and more significant. In sum, the results of the estimation are not driven by the inclusion of the SEO sample. However, some findings become more ambiguous after the SEO sample is omitted.

2.3.3 NLSY79-CH Results

Section 2.3.3 contains the basic estimates for the NLSY79-CH sample. Section 2.3.3 discusses additional items from the NLSY79-CH data. Section 2.3.3 performs some robustness checks on the NLSY79-CH results.

Basic Estimates

In order to elucidate a possible mechanism behind the observed influence of childhood economic conditions on adult outcomes, this section examines how the quality of a child's home environment varies over the business cycle. Table 2.9 exhibits the impact of the current state or national unemployment rate on the standardized values of the total score as well as the cognitive stimulation and emotional support subscores from the HOME-SF inventory.[37] Four sets of esti-

[37] The scores for each part of the home inventory are standardized among all members of the sample at the same age level.

mates are computed, one for each stage of development: infant/toddler, early childhood, middle childhood, and early adolescence. The specifications using the national unemployment rate as a regressor control for race, gender, state of residence, indicator variables for age, and a linear trend in survey year.[38] The specifications containing the state unemployment rate as a regressor control for race, gender, state of residence, indicator variables for age, fixed effects for survey year, and a linear trend in survey year specific to each state of residence.[39] In most cases, the point estimate for the coefficient on the unemployment rate is negative. The national unemployment rate has a significantly negative impact on the emotional score in early adolescence. The state unemployment rate has a significantly negative impact on the total and emotional scores in middle childhood as well as the total score in early adolescence.

I now assemble several pieces of evidence to determine whether the impact of the unemployment rate on the home environment is likely to be due to variation over the business cycle in the characteristics of parents raising children or to a causal effect of macroeconomic conditions on the caregiving behavior of parents. I first examine how the estimates for the specifications in Table 2.9 change from before to after controlling for maternal background variables such as year born, education, occupation, and test score.[40] Table 2.10 reports the findings from this comparison. On the whole, the estimated coefficient on the unemployment rate does not change drastically as a result of including maternal background variables. Although the addition of these variables sometimes attenuates the negative coefficient on the unemployment rate, the negative impact of the state unemployment rate on the cognitive score in early adolescence gains significance after controlling for them.

I next analyze the relationship between the current unemployment rate and the characteristics of mothers raising children. To perform this exercise, each observation on a child from Table 2.10 is matched to data on the schooling level and test score of the child's mother. Indicators for the mother's test performance and education level are then regressed on the current national or state unemployment rate.[41] The specifications using the national unemployment rate as a regressor con-

[38]The standard errors are clustered by survey year.

[39]The standard errors are clustered by state of residence.

[40]See the note to Table 2.10 for more details on the sample selection criteria as well as the coding of the maternal background variables. Recall that the children in Table 2.10 have a mother who is a participant in the NLSY79.

[41]The test score used here is the Armed Forces Qualifying Test. During the summer and fall of 1980, the Armed

Table 2.9: Relationship of Current Unemployment Rate to HOME-SF Inventory Scores for NLSY79-CH Sample

	HOME-SF Inventory		
	Standardized Total Score	Standardized Cognitive Score	Standardized Emotional Score
		Part A: Infant/Toddler	
		Current National Unemployment Rate	
U.E. Rate	-0.0528 (0.0655)	-0.0423 (0.0629)	-0.0456 (0.0557)
		Current State Unemployment Rate	
U.E. Rate	-0.0078 (0.0165)	-0.0026 (0.0168)	-0.0097 (0.0155)
Sample Size	12 Years / 50 States / 5,280 Individuals / 6,505 Observations		
		Part B: Early Childhood	
		Current National Unemployment Rate	
U.E. Rate	0.0017 (0.0528)	-0.0022 (0.0403)	0.0034 (0.0530)
		Current State Unemployment Rate	
U.E. Rate	-0.0111 (0.0176)	-0.0135 (0.0143)	-0.0061 (0.0177)
Sample Size	12 Years / 50 States / 6,386 Individuals / 8,187 Observations		
		Part C: Middle Childhood	
		Current National Unemployment Rate	
U.E. Rate	-0.0191 (0.0514)	-0.0180 (0.0477)	-0.0130 (0.0397)
		Current State Unemployment Rate	
U.E. Rate	-0.0332 (0.0115)	-0.0136 (0.0168)	-0.0446 (0.0146)
Sample Size	12 Years / 50 States / 7,523 Individuals / 11,749 Observations		
		Part D: Early Adolescence	
		Current National Unemployment Rate	
U.E. Rate	-0.0466 (0.0277)	-0.0140 (0.0339)	-0.0663 (0.0243)
		Current State Unemployment Rate	
U.E. Rate	-0.0550 (0.0236)	-0.0334 (0.0173)	-0.0581 (0.0314)
Sample Size	11 Years / 50 States / 6,620 Individuals / 11,259 Observations		

Note: The main estimation sample for the NLSY79-CH is used to generate the estimates above. The scores for each part of the HOME-SF inventory are standardized among all individuals at the same age level in the sample. The upper set of estimates for each part of the HOME-SF inventory control for race, gender, state of residence, a linear trend in survey year, and fixed effects for age at the end of the survey year. The lower set of estimates for each part of the HOME-SF inventory control for race, gender, state of residence, indicator variables for survey year, fixed effects for age at the end of the survey year, and a linear trend in survey year specific to each state of residence. Huber-White standard errors, clustered by survey year for the upper set of estimates in each part and by state of residence for the lower set of estimates in each part, are reported in parentheses.

Table 2.10: Relationship of Current Unemployment Rate to HOME-SF Inventory Scores for NLSY79-CH Sample Before and After Controlling for Maternal Background Variables

	HOME-SF Inventory					
	Standardized Total Score		Standardized Cognitive Score		Standardized Emotional Score	
	Part A: Infant/Toddler					
	Current National Unemployment Rate					
U.E. Rate	-0.0519 (0.0643)	-0.0534 (0.0594)	-0.0404 (0.0626)	-0.0410 (0.0581)	-0.0464 (0.0551)	-0.0482 (0.0525)
	Current State Unemployment Rate					
U.E. Rate	-0.0038 (0.0162)	-0.0060 (0.0147)	0.0028 (0.0173)	0.0008 (0.0162)	-0.0094 (0.0154)	-0.0113 (0.0151)
Maternal Background	No	Yes	No	Yes	No	Yes
Sample Size	12 Years / 50 States / 5,051 Individuals / 6,232 Observations					
	Part B: Early Childhood					
	Current National Unemployment Rate					
U.E. Rate	-0.0048 (0.0513)	-0.0030 (0.0417)	-0.0068 (0.0392)	-0.0062 (0.0291)	-0.0033 (0.0514)	-0.0009 (0.0457)
	Current State Unemployment Rate					
U.E. Rate	-0.0076 (0.0164)	-0.0042 (0.0139)	-0.0097 (0.0131)	-0.0059 (0.0139)	-0.0044 (0.0167)	-0.0007 (0.0150)
Maternal Background	No	Yes	No	Yes	No	Yes
Sample Size	12 Years / 50 States / 6,105 Individuals / 7,849 Observations					
	Part C: Middle Childhood					
	Current National Unemployment Rate					
U.E. Rate	-0.0185 (0.0518)	-0.0143 (0.0427)	-0.0138 (0.0502)	-0.0095 (0.0427)	-0.0169 (0.0377)	-0.0145 (0.0309)
	Current State Unemployment Rate					
U.E. Rate	-0.0267 (0.0102)	-0.0172 (0.0086)	-0.0019 (0.0151)	0.0087 (0.0112)	-0.0469 (0.0156)	-0.0423 (0.0160)
Maternal Background	No	Yes	No	Yes	No	Yes
Sample Size	12 Years / 50 States / 7,195 Individuals / 11,270 Observations					
	Part D: Early Adolescence					
	Current National Unemployment Rate					
U.E. Rate	-0.0498 (0.0271)	-0.0419 (0.0247)	-0.0150 (0.0335)	-0.0067 (0.0322)	-0.0708 (0.0234)	-0.0664 (0.0206)
	Current State Unemployment Rate					
U.E. Rate	-0.0536 (0.0236)	-0.0603 (0.0218)	-0.0259 (0.0172)	-0.0324 (0.0143)	-0.0641 (0.0313)	-0.0684 (0.0306)
Maternal Background	No	Yes	No	Yes	No	Yes
Sample Size	11 Years / 50 States / 6,339 Individuals / 10,815 Observations					

Note: The dataset used here is constructed by restricting the main estimation sample for the NLSY79-CH to observations on individuals whose mother has information on first occupation, years of schooling, and AFQT score. The scores for each part of the HOME-SF inventory are standardized among all individuals at the same age level in the sample. Maternal background variables are indicators for mother's first occupation, educational attainment, AFQT quartile, and birth year. The quartiles for the AFQT score are computed by comparing each mother's AFQT score to the AFQT scores of all female respondents in the NLSY79 with the same year of birth. The categories for occupation after first leaving school are the 23 major occupational groups in the 2000 SOC. The categories for educational attainment are: less than high school graduate, high school diploma, some college, college degree, and some post-graduate training. The upper set of estimates for each part of the HOME-SF inventory control for race, gender, state of residence, a linear trend in survey year, and fixed effects for age at the end of the survey year. The lower set of estimates for each part of the HOME-SF inventory control for race, gender, state of residence, indicator variables for survey year, fixed effects for age at the end of the survey year, and a linear trend in survey year specific to each state of residence. Huber-White standard errors, clustered by survey year for the upper set of estimates in each part and by state of residence for the lower set of estimates in each part, are reported in parentheses.

trol for mother's race, state of residence, fixed effects for mother's year of birth, and a linear trend in survey year. The specifications using the state unemployment rate as a regressor control for mother's race, state of residence, fixed effects for mother's year of birth, indicator variables for survey year, and a linear trend in survey year specific to each state of residence. Table 2.11 displays the results of the analysis. Although the coefficient on the unemployment rate is statistically insignificant in most cases, the unemployment rate does have a significantly negative relationship with some measures of parental quality. Hence, it is possible that the observed negative impacts of the unemployment rate on the home environment could be attributed in part to changes over the business cycle in the characteristics of parents raising children.

To reduce the potential influence of underlying changes in parental quality on the results, I compute family and person fixed-effects estimates for the impact of the national unemployment rate on the home environment. The former and latter sets of estimates are presented in Tables 2.12 and 2.13, respectively.[42] The specifications contain indicator variables for age and a linear trend in survey year.[43] The point estimate for the coefficient on the unemployment rate is negative in most cases. In the family fixed-effects regressions, the unemployment rate has a significantly negative impact on the total, cognitive, and emotional scores for infants/toddlers and early adolescents. In the person fixed-effects regressions, the unemployment rate has a significantly negative impact on the cognitive score in middle childhood as well as the total, cognitive, and emotional scores in early adolescence. Overall, the findings from the family and person fixed-effects regressions are consistent with a causal effect of macroeconomic conditions on parental caregiving behavior.

Additional Items

To investigate the factors contributing to a relationship between the unemployment rate and the quality of a child's home environment, several items from each part of the HOME-SF inven-

---

Services Vocational Aptitude Battery was administered to participants in the NLSY79. In the regressions, the indicators for mother's test performance are binary variables for the mother's AFQT score being above the first, second, and third quartiles of all female participants in the NLSY79 with the same year of birth.

[42]To construct the dataset in Table 2.12, the sample from each part of Table 2.9 is limited to observations on individuals having a sibling who belongs to the sample from that part of Table 2.9 in some survey year. To construct the dataset in Table 2.13, the sample from each part of Table 2.9 is limited to observations on individuals who appear in the sample from that part of Table 2.9 in at least two survey years.

[43]A gender dummy is also included in the family fixed-effects regressions.

Table 2.11: Relationship of Unemployment Rates in Childhood to Maternal AFQT and Schooling

	AFQT $\geq$ Q1	AFQT $\geq$ Q2	AFQT $\geq$ Q3	H.S. Diploma	College Degree	Grad. School
			Mothers of Individuals in Home Inventory Part A			
			National Unemployment Rate for Youth when Infant/Toddler			
U.E. Rate	-0.0024 (0.0046)	0.0054 (0.0053)	0.0005 (0.0045)	0.0026 (0.0101)	-0.0005 (0.0047)	0.0011 (0.0013)
			State Unemployment Rate for Youth when Infant/Toddler			
U.E. Rate	0.0005 (0.0069)	-0.0096 (0.0088)	-0.0069 (0.0070)	-0.0096 (0.0079)	-0.0106 (0.0082)	0.0004 (0.0035)
Sample Size			12 Years / 50 States / 6,232 Observations			
			Mothers of Individuals in Home Inventory Part B			
			National Unemployment Rate for Youth in Early Childhood			
U.E. Rate	0.0040 (0.0063)	-0.0043 (0.0064)	-0.0056 (0.0054)	0.0023 (0.0155)	-0.0103 (0.0047)	-0.0019 (0.0016)
			State Unemployment Rate for Youth in Early Childhood			
U.E. Rate	-0.0138 (0.0097)	-0.0126 (0.0106)	-0.0120 (0.0095)	-0.0215 (0.0059)	-0.0060 (0.0065)	0.0003 (0.0028)
Sample Size			12 Years / 50 States / 7,849 Observations			
			Mothers of Individuals in Home Inventory Part C			
			National Unemployment Rate for Youth in Middle Childhood			
U.E. Rate	-0.0062 (0.0112)	-0.0074 (0.0076)	-0.0029 (0.0052)	-0.0025 (0.0213)	-0.0054 (0.0037)	-0.0007 (0.0016)
			State Unemployment Rate for Youth in Middle Childhood			
U.E. Rate	-0.0085 (0.0059)	-0.0074 (0.0060)	-0.0051 (0.0061)	-0.0165 (0.0048)	-0.0069 (0.0041)	-0.0003 (0.0023)
Sample Size			12 Years / 50 States / 11,270 Observations			
			Mothers of Individuals in Home Inventory Part D			
			National Unemployment Rate for Youth in Early Adolescence			
U.E. Rate	-0.0045 (0.0034)	0.0021 (0.0031)	0.0033 (0.0031)	-0.0170 (0.0111)	0.0007 (0.0038)	0.0010 (0.0013)
			State Unemployment Rate for Youth in Early Adolescence			
U.E. Rate	0.0068 (0.0097)	-0.0058 (0.0086)	0.0014 (0.0067)	0.0090 (0.0068)	-0.0004 (0.0055)	-0.0042 (0.0018)
Sample Size			11 Years / 50 States / 10,815 Observations			

Note: The dataset used here is constructed as follows. First, the main estimation sample for the NLSY79-CH is restricted to individuals whose mother has information on first occupation, years of schooling, and AFQT score. Second, every observation on an individual in the resulting sample is matched to the available information on the individual's mother. The quartiles for the AFQT score are computed by comparing each mother's AFQT score to the AFQT scores of all female respondents in the NLSY79 with the same year of birth. The upper set of estimates for each part of the HOME-SF inventory control for mother's race, state of residence, fixed effects for mother's year of birth, and a linear trend in survey year. The lower set of estimates for each part of the HOME-SF inventory control for mother's race, state of residence, fixed effects for mother's year of birth, indicator variables for survey year, and a linear trend in survey year specific to each state of residence. Huber-White standard errors, clustered by survey year for the upper set of estimates in each part and by state of residence for the lower set of estimates in each part, are reported in parentheses.

Table 2.12: Family Fixed-Effects Estimates for Relationship of Current Unemployment Rate to HOME-SF Inventory Scores for NLSY79-CH Sample

	HOME-SF Inventory		
	Standardized Total Score	Standardized Cognitive Score	Standardized Emotional Score
		Part A: Infant/Toddler	
		Current National Unemployment Rate	
U.E. Rate	-0.0565 (0.0201)	-0.0437 (0.0208)	-0.0485 (0.0220)
Sample Size	1,576 Families / 3,779 Individuals / 4,723 Observations		
		Part B: Early Childhood	
		Current National Unemployment Rate	
U.E. Rate	-0.0003 (0.0136)	-0.0059 (0.0149)	0.0035 (0.0160)
Sample Size	1,988 Families / 4,899 Individuals / 6,393 Observations		
		Part C: Middle Childhood	
		Current National Unemployment Rate	
U.E. Rate	-0.0142 (0.0107)	-0.0197 (0.0108)	-0.0017 (0.0130)
Sample Size	2,449 Families / 6,343 Individuals / 10,080 Observations		
		Part D: Early Adolescence	
		Current National Unemployment Rate	
U.E. Rate	-0.0527 (0.0120)	-0.0293 (0.0116)	-0.0590 (0.0143)
Sample Size	2,163 Families / 5,539 Individuals / 9,604 Observations		

Note: The dataset used here is constructed by restricting the main estimation sample for each part of the HOME-SF inventory to observations on individuals who have a sibling in the sample for that part of the HOME-SF inventory in some survey year. The scores for each part of the HOME-SF inventory are standardized among all individuals at the same age level in the sample. The specifications include a linear trend in survey year and fixed effects for age at the end of each survey year. Huber-White standard errors, clustered by family, are reported in parentheses.

Table 2.13: Person Fixed-Effects Estimates for Relationship of Current Unemployment Rate to HOME-SF Inventory Scores for NLSY79-CH Sample

		HOME-SF Inventory	
	Standardized Total Score	Standardized Cognitive Score	Standardized Emotional Score
		Part A: Infant/Toddler	
		Current National Unemployment Rate	
U.E. Rate	-0.0663 (0.0352)	-0.0643 (0.0361)	-0.0346 (0.0382)
Sample Size		1,225 Individuals / 2,450 Observations	
		Part B: Early Childhood	
		Current National Unemployment Rate	
U.E. Rate	-0.0173 (0.0215)	-0.0329 (0.0244)	0.0029 (0.0263)
Sample Size		1,801 Individuals / 3,602 Observations	
		Part C: Middle Childhood	
		Current National Unemployment Rate	
U.E. Rate	-0.0272 (0.0146)	-0.0425 (0.0151)	0.0017 (0.0183)
Sample Size		4,080 Individuals / 8,306 Observations	
		Part D: Early Adolescence	
		Current National Unemployment Rate	
U.E. Rate	-0.0451 (0.0149)	-0.0304 (0.0149)	-0.0442 (0.0176)
Sample Size		3,924 Individuals / 8,563 Observations	

Note: The dataset used here is constructed by restricting the main estimation sample for each part of the HOME-SF inventory to observations on individuals who are in the sample for that part of the HOME-SF inventory in at least two years. The scores for each part of the HOME-SF inventory are standardized among all individuals at the same age level in the sample. The specifications include a linear trend in survey year and fixed effects for age at the end of each survey year. Huber-White standard errors, clustered by person, are reported in parentheses.

tory are separately analyzed.[44] The specifications from Tables 2.9, 2.10, 2.11, 2.12, and 2.13 are reestimated using the individual items instead of the aggregate scores as dependent variables. The outcomes examined from each part of the HOME-SF inventory are as follows with the applicable parts listed in parentheses: an indicator for the child having at least one children's book (A, B, C, D); an indicator for the mother reading to the child at least once a week (A, B, C); an indicator for the mother taking the child to the grocery at least once a week (A); the number of cuddly or role-playing toys that the child has (A); the number of push or pull toys that the child has (A); an indicator for the child seeing his/her father daily (A, B); an indicator for the child eating with both his/her mother and father at least once a day (A, B, C, D); an indicator for the mother having spanked the child in the past week (A, B, C, D); an indicator for the child's family receiving at least one magazine regularly (B); an indicator for the child having a tape recorder or similar device (B); an indicator for the child having been taken to a museum in the past year (B, C, D); the number of hours that the television is on at home (B); an indicator for the child having a musical instrument to use at home (C, D); an indicator for the child's family receiving a daily newspaper (C, D); an indicator for the child being taken to the theater in the past year (C, D); an indicator for the child participating in special lessons or extracurricular activities (C, D); an indicator for the child spending time daily with his/her father (C, D); an indicator for the parents discussing television programs with the child (C, D).

In regressions similar to those from Table 2.9, the unemployment rate is seen to have different influences at different stages of childhood. Using the national unemployment rate as a regressor, the only statistically significant result is a negative impact in early adolescence on the probability of having been taken to a museum in the past year. Using the state unemployment rate as a regressor, a significant negative impact is found on: the number of cuddly or role-playing toys for infants/toddlers; the presence of a tape recorder at home in early childhood; the probabilities in middle childhood of spending time daily with one's father and of eating with both one's mother and father at least once a day; the probabilities in early adolescence of having access to a musical instrument at home, of having been taken to a museum in the past year, of eating with both one's mother and father at least once a day, and of having been spanked by one's mother in the past week. The state unemployment rate is also seen to have a significant positive impact in early childhood

[44]The estimates summarized here are available from the author on request.

on the probability of having been taken to a museum in the past year.

To assess whether changes over the business cycle in the characteristics of parents raising children are affecting the results, the regressions are performed both including and excluding maternal background variables as in Table 2.10. In most cases, the estimates do not change substantially depending on whether or not maternal background variables are added to the specifications. However, the national unemployment rate has a significant negative impact on participation in special lessons among early adolescents if and only if maternal background controls are excluded, and the state unemployment rate has a significant negative impact in early adolescence on the probabilities of having been taken to a museum in the past year and of having been spanked by one's mother in the past week if and only if maternal background controls are included.[45]

In addition, family and person fixed-effects estimates are computed as in Tables 2.12 and 2.13, in order to account for a possible relationship between the unemployment rate and parental background. In the family fixed-effects regressions, a significant negative impact of the national unemployment rate is found on: the probability of the child's family receiving at least one magazine regularly in early childhood; the probabilities in middle childhood of having access to a musical instrument at home, of having been taken to a museum in the past year, and of having been taken to the theater in the past year; the probabilities in early adolescence of having been taken to a museum in the past year and of participating in special lessons. The unemployment rate also has a significant positive impact in these regressions on the probability among early adolescents of being spanked by one's mother in the past week. In the person fixed-effects regressions, a significant positive impact of the national unemployment rate is found on the probabilities in middle childhood of having been taken to a museum in the past year, of having been taken to the theater in the past year, and of participating in special lessons as well as the probability in early adolescence of having been taken to a museum in the past year. The unemployment rate also has a significant positive impact in these regressions on the probability among infants/toddlers of having been spanked by one's mother in the past week.

The discussion thus far has paid little attention to parental behavior during the prenatal and postnatal period. However, the NLSY79-CH contains a number of questions related to this topic,

[45] Recall that the sample from Table 2.10 is a subset of the sample from Table 2.9. Hence, the estimates for the specification without maternal background variables using the sample from Table 2.10 can differ from those using the sample from Table 2.9.

which I use to document the impact of economic conditions on prenatal and postnatal care.[46] The outcomes examined are: an indicator for the mother visiting a medical professional for prenatal care; indicators for the mother drinking alcohol and smoking in the twelve months before the child's birth; indicators for the mother taking vitamins, reducing caloric intake, and lowering salt consumption during pregnancy; the gestation length for the child in weeks; indicators for the child being born preterm, at term, and postterm; an indicator for the child being delivered by Caesarean section; the child's birth weight in ounces; indicators for the child being born with low birth weight, very low birth weight, and extremely low birth weight; the child's length at birth in inches; indicators for the child being ever breastfed, breastfed for at least six months, breastfed for at least one year, and breastfed for at least eighteen months.[47]

The main estimation sample consists of children who have valid data on the variables listed above as well as information on year of birth and mother's state of residence at age fourteen.[48] Only individuals whose mother belongs to the cross-sectional or supplemental sample of the NLSY79 are included. The analysis focuses on macroeconomic conditions around the time of a child's birth. The average unemployment rate in the year before the child's birth and the year of the child's birth is used for outcomes related to the period before or at birth, and the average unemployment rate in the year of the child's birth and the year after the child's birth is used for outcomes related to the period after birth.[49]

I begin by performing an analysis similar to that in Table 2.9. In specifications using the national unemployment rate around the time of the child's birth as a regressor, the control variables are indicators for race and gender, fixed effects for mother's state of residence at age fourteen, and a

---

[46]The results summarized here are available from the author on request.

[47]A birth is defined as preterm if the gestation length is 36 weeks or less, at term if the gestation length is between 37 and 42 weeks inclusive, and postterm if the gestation length is 43 weeks or more. An infant's birth weight is defined as low if it is no more than 88 ounces, very low if it is no more than 52 ounces, and extremely low if it is no more than 35 ounces.

[48]The dataset is limited to individuals whose mother was residing in one of the fifty states or the District of Columbia at age fourteen.

[49]The outcomes related to the period before or at birth are: the mother visiting a medical professional for prenatal care; the mother smoking and drinking during pregnancy; the mother taking vitamins, reducing calories, lowering salt; the child's gestation length; the mother giving birth pre-term, at term, and postterm; the mother having a Caesarean section; the child's birth weight; the child's birth weight being low, very low, and extremely low; the child's length at birth. The outcomes related to the period after birth are the child being ever breastfed as well as the child being breastfed for at least six months, one year, and eighteen months.

linear trend in year of birth.[50] In specifications using the state unemployment rate around the time of the child's birth as a regressor, the control variables are indicators for race and gender, fixed effects for mother's state of residence at age fourteen and for child's year of birth, and a linear trend in child's year of birth specific to each state of residence for the mother at age fourteen.[51] The main findings from these regressions are as follows. A significant positive impact of the national unemployment rate around birth is found on the probabilities of the mother lowering salt consumption during pregnancy, of the child ever being breastfed, and of the child being breastfed for at least six months. A significant negative impact of the state unemployment rate around birth is found on the probability of the mother drinking alcohol in the twelve months before the child's birth as well as the child's gestation length and length at birth.

I also conduct several exercises to determine whether the results are driven by a relationship between the unemployment rate and parental background. As in Table 2.10, I estimate the specifications including and excluding maternal background variables. In general, the main findings do not change much with the addition of these variables. As in Table 2.11, I test for an association of the unemployment rate with the schooling levels and test scores of parents. In specifications using the national unemployment rate around the time of the child's birth as a regressor, the control variables are dummies for child's race, fixed effects for mother's state of residence at age fourteen and for mother's year of birth, and a linear trend in child's year of birth. In specifications using the state unemployment rate around the time of the child's birth as a regressor, the control variables are dummies for child's race, fixed effects for mother's state of residence at age fourteen and for mother's year of birth, indicators for child's year of birth, and a linear trend in child's year of birth specific to each state of residence for the mother at age fourteen. These regressions indicate that the unemployment rate has a significant positive impact on the probability of the child's mother having completed high school. Hence, it is possible that changes over the business cycle in the characteristics of mothers giving birth could explain improved prenatal and postnatal care during a recession.

To lessen the potential influence of parental background on the results, I compute family fixed-effects estimates as in Table 2.12 using children with a sibling also in the sample. The specification

[50] The standard errors are clustered by year born.

[51] The standard errors are clustered by mother's state of residence at age fourteen.

controls for gender as well as a linear trend in year of birth. In these regressions, the national unemployment rate is seen to have a significant positive impact on the probabilities of the mother reducing caloric intake during pregnancy, of the mother lowering salt consumption during pregnancy, of the child ever being breastfed, and of the child being breastfed for six months. Overall, a higher unemployment rate appears to generate an improvement in some aspects of prenatal and postnatal care even after accounting for the quality of mothers giving birth.

Robustness Checks

This section describes the results from a series of robustness checks parallel to those in sections 2.3.1 and 2.3.2. First, the specifications in the lower row of each panel from Tables 2.9, 2.10, and 2.11 are reestimated using the rate of insured unemployment for each state instead of the state unemployment rate as a regressor. This substitution does not substantially affect the conclusions of the analysis. However, state unemployment is now seen to have a significantly negative impact in Table 2.9 on the emotional score for early adolescents. In addition, the negative impact of state unemployment in Table 2.10 on the cognitive score for early adolescents does not gain significance after the inclusion of parental background variables, and the negative relationship of state unemployment to maternal high school completion in Table 2.11 is no longer statistically significant for individuals in middle childhood.

Second, the estimates in Tables 2.9, 2.10, 2.11, 2.12, and 2.13 are replicated using the employment-to-population ratio instead of the unemployment rate as a gauge of labor market tightness. On balance, the findings from Table 2.9 do not change much as a result of this modification. Worse national economic conditions are now seen to have a significant negative impact on the total score for early adolescents, and worse state economic conditions are now seen to have a significant negative impact on the total and cognitive scores in early childhood. Nonetheless, the negative impacts of worse state economic conditions on the total and emotional scores in middle childhood as well as the total score in early adolescence are no longer statistically significant.

In Table 2.10, the results using the employment-to-population ratio usually resemble those using the unemployment rate, although some changes should be noted. Worse national economic conditions are observed to have a significant negative impact on the total score for early adolescents both before and after controlling for maternal background variables. By contrast, the negative impacts of worse state economic conditions on total and cognitive scores in early childhood lose significance after maternal background variables are added to the specification, and no statistically significant impact of state economic conditions on the total, emotional, or cognitive score is detected in middle childhood or early adolescence irrespective of the inclusion of maternal background variables. In Table 2.11, there continues to be some evidence of a significant relationship between economic conditions and maternal characteristics.

Computing family and person fixed-effects estimates as in Tables 2.12 and 2.13, a significant impact of economic conditions on the home environment is detected at several stages of childhood when using the employment-to-population ratio instead of the unemployment rate as a regressor. In addition to the significant results originally obtained in Table 2.12, a significant negative impact of worse economic conditions is found on the total and cognitive scores in early childhood as well as the total, cognitive, and emotional scores in middle childhood. In addition to the significant results originally obtained in Table 2.13, a significant negative impact of worse economic conditions is found on the total and cognitive scores for infants/toddlers as well as the total score in middle childhood.

Third, the estimates in Tables 2.9, 2.10, 2.11, 2.12, and 2.13 are recalculated using only children with a mother in the cross-sectional sample of the NLSY79.[52] This restriction leaves the basic pattern of results in Table 2.9 mostly unaffected. However, the negative impacts of the state unemployment rate on the total scores in middle childhood and early adolescence lose significance.

On balance, the findings from Table 2.10 do not change drastically as a result of excluding children with a mother in the supplemental sample. The national unemployment rate now has a significant negative impact on the total score in early adolescence both before and after the addition of maternal background variables, but the state unemployment rate no longer has a significant negative impact on the total score in middle childhood before or after maternal background variables are included. In addition, the negative impact of the state unemployment rate on the total score in early adolescence is now significant only after controlling for maternal background, and the negative impact of the state unemployment rate on the cognitive score in early adolescence no longer gains significance after the inclusion of maternal background variables.

[52]The original dataset contains individuals with a mother in the cross-sectional or supplemental sample of the NLSY79. Blacks, hispanics, and disadvantaged whites are overrepresented in the supplemental sample.

In Table 2.11, there is only weak evidence of a systematic relationship between the unemployment rate and maternal characteristics once the analysis is limited to children whose mother belongs to the cross-sectional sample. In particular, the national unemployment rate has a significant negative impact only on the probability for early adolescents of one's mother completing high school and a significant positive impact only on the probability for early adolescents of one's mother having an AFQT score above the second quartile. Using the restricted sample, the state unemployment rate does not have a significant impact on any of the maternal characteristics examined.

The findings from the family and person fixed-effects regressions in Tables 2.12 and 2.13 are largely robust to the removal of the supplemental sample. The only notable change is that the negative impact of the national unemployment rate on the emotional score for infants/toddlers is no longer statistically significant in the family fixed-effects regressions.

2.4 Conclusion

This paper employs several strategies to examine the influence of macroeconomic conditions in childhood on labor market outcomes as an adult. Using a large sample of individuals from the Census, I find significant evidence of a negative relationship between the average unemployment rate in childhood and some measures of schooling, employment, and income. However, the estimates provide little support for the hypothesis that economic conditions earlier versus later in childhood are a bigger determinant of adult outcomes.

Using a matched sample of parents and children from the PSID, I demonstrate that the findings cannot be easily attributed to selection by parents over the business cycle into raising children. In particular, the impact of childhood economic conditions does not weaken much after controlling for parental background variables, and the quality of parents raising children does not appear to be lower during an economic downturn. Moreover, a negative impact of the unemployment rate in childhood on adult economic performance is observed in family fixed-effects regressions, although the estimates are not always statistically significant. The results are typically robust to changes in the construction of the sample and the measurement of economic conditions.

Using detailed information on parental caregiving behavior from the NLSY79-CH, I investigate how the quality of a child's home environment varies over the business cycle, in order to illustrate a possible mechanism behind the impact of childhood economic conditions. The data indicate that some measures of a child's household surroundings deteriorate during a recession. Family and person fixed-effects estimates confirm that the negative influence of the unemployment rate on parental caregiving cannot be entirely explained by a relationship between macroeconomic conditions and the background characteristics of parents raising children.

In sum, the evidence in this paper is consistent with a causal effect of childhood economic conditions on parental investments in children as well as the stock of human capital in adulthood. The observed impacts are often large in magnitude. However, the effect sizes can vary substantially with the sample used, the outcome analyzed, and the estimation strategy employed. In terms of policy implications, the empirical results provide a possible rationale for targeting economic stimulus programs towards children. Policies designed to improve a child's home environment might help mitigate some of the adverse impacts of a recession on adult economic outcomes. Improvements in neighborhoods and schools might also be beneficial for this purpose, although these mechanisms were not explored in this paper. Additionally, the empirical results do not seem to suggest a strong advantage to concentrating assistance on younger versus older children.

Chapter 3

Sequential Exchange with Stochastic Transaction Costs

This chapter is joint work with Yuichiro Kamada.

3.1 Introduction

This paper is motivated by the following situation. Two firms operating in different markets find it potentially profitable to exchange trade secrets, but there is a cost for transferring knowledge from one firm to the other. This cost might represent the resources spent training employees of the other firm or the expense of encrypting data to protect secrets from outsiders. Because of the intangible nature of information, it is infeasible for the two parties to write a court-enforceable contract specifying the goods to be traded. Moreover, if one party immediately reveals all of its information to the other party, then the latter would have no incentive to reveal its information to the former, because transferring knowledge is costly. In this situation, how can the two parties exchange their knowledge with each other?

We study the bilateral exchange of divisible goods in a continuous-time environment. Each of two agents possesses an equal amount of a different good that only the other agent values. In order to transfer some amount of her own good to the other party, an agent must pay a transaction cost that evolves according to a geometric Brownian motion.[1] Although the first-best solution requires each agent to make a single transfer of all her good to the other agent once the cost reaches a certain cutoff value, such a policy cannot be supported as part of a subgame-perfect equilibrium in the absence of court-enforceable contracts, because each agent is unwilling to transfer her own good after receiving all the other agent's good.

Our main objective is to solve for an incentive-compatible transaction scheme in which agents can realize positive gains from trade by making a sequence of gradually decreasing transfers. We obtain a closed-form solution for the unique maximal symmetric subgame-perfect equilibrium. Due to the uniqueness of the result, we can derive a number of meaningful comparative statics. Specifically, we show that as agents become infinitely patient, the expected discounted payoff from the second-best policy converges to the efficient outcome, provided that the drift of the cost process is not excessively high relative to its volatility.

The idea behind the construction of an incentive-compatible scheme with positive transfers is simple. In equilibrium, when the two agents make transfers, they withhold some amount of the goods. The agents continue to transfer the withheld goods if and only if both parties have made the prescribed transfers in the past. Because the agents receive a positive continuation value if and only if they make the prescribed transfers, they are willing to incur a positive transaction cost in equilibrium.

The key tradeoff lies between *the size of the next transfer* and *the waiting time until the transfer*. In order for agents to anticipate a high ex ante payoff, they must exchange a large amount of the goods at the next transaction. Hence, the amount withheld must be small. Because the resulting continuation value is low, the agents are willing to incur only a small cost when making the transaction. However, the expected waiting time for a small cost to realize is high. Since agents discount the future, a lengthy waiting time reduces the ex ante expected payoff. The maximal symmetric equilibrium derived in this paper balances the tradeoff between the size of the next transfer and the waiting time until the transfer.

Methodologically, we combine the theory of optimal investment under uncertainty with the the-

[1] The assumption of geometric Brownian motion is standard in the literature on investment under uncertainty. For example, geometric Brownian motion is used by McDonald and Siegel (1986) to model the market value of an investment, by Dixit (1991) to model a demand parameter, and by Bertola and Caballero (1994) to model an index of business conditions.

ory of repeated games by imposing incentive-compatibility constraints on investment decisions.[2] The difference from standard models of repeated games is that the set of feasible actions and hence the set of feasible instantaneous payoffs irreversibly changes depending on the actions taken by agents. This feature makes it difficult to extrapolate from the literature on stochastic games, because existing research generally assumes irreducibility of the state transition.[3]

Besides the aforementioned example, our framework fits a number of social situations in the real world where exchanges of goods are involved. For instance, consider the exchange of prisoners or hostages between two groups that are enemies of each other or that are at war, as in the transfer of captives that took place between the Taliban and the North Alliance.[4] The two agents in the model would be representatives of the parties who are negotiating for the release of prisoners. Each group typically does not derive any benefit from holding prisoners of war from the other group, but each group wishes to exchange these captives for the return of its own prisoners held by the other group. The stochastic transfer cost might be a measure of the political climate or degree of tensions between the two groups. When the tensions are high and the relationship is more hostile, each group has a high cost of releasing prisoners from the other group. When the tensions are low and the relationship is less hostile, each group has a low cost of releasing prisoners from the other group. In the example of the Taliban and the Northern Alliance, what happened was a gradual exchange of prisoners, as our model predicts.

A further prediction of our model is that the size of transfers decreases over time. An example of this pattern in the real world is provided in Article Two of the Treaty with the Creeks at Indian Springs in February of 1812.[5] The following extract describes how the United States intended to compensate the Creek Indians for lands ceded to the state of Georgia:

> But whereas said Creek nation have considerable improvements within the limits of the territory hereby ceded, and will, moreover, have to incur expenses in their removal, it is further stipulated that, for the purpose of rendering a fair equivalent for the losses and inconveniences which said nation will sustain by removal, and to enable them to obtain supplies in their new

[2] See footnote 1 for references to the literature on investment under uncertainty.

[3] See Dutta (1995), Fudenberg and Yamamoto (2011), and Horner et al. (2011) for folk theorems in stochastic games at varying levels of generality.

[4] A news article on prisoner exchanges between the Taliban and the Northern Alliance can be found at `http://news.bbc.co.uk/2/hi/south_asia/203926.stm`.

[5] A copy of this document can be found at `http://georgiainfo.galileo.usg.edu/indspri2.htm`.

> settlement, the United States agree to pay to the nation emigrating from the lands herein ceded the sum of four hundred thousand dollars; of which amount there shall be paid to said party of the second part, as soon as practicable after the ratification of this treaty, the sum of two hundred thousand dollars. And as soon as the said party of the second part shall notify the Government of the United States of their readiness to commence their removal, there shall be paid the further sum of one hundred thousand dollars. And the first year after said emigrating party shall have settled in their new country, they shall receive, of the amount first above named, the further sum of twenty-five thousand dollars; and the second year, the sum of twenty-five thousand dollars; and annually thereafter, the sum of five thousand dollars, until the whole is paid.

According to the agreement, the United States would make a gradually decreasing series of monetary payments to the Creek Indians as they progressively surrendered their existing lands and relocated to a different territory.

The remainder of this paper is organized as follows. Section 3.2 reviews the related literature. Section 3.3 outlines our basic model of bilateral trade with stochastic transaction costs and defines the concept of a maximal symmetric subgame-perfect equilibrium in the context of our model. In section 3.4, we analyze the model in several steps. First, we prove an impossibility result stating that if (a) the transaction cost is bounded below by a positive number and (b) the quantity of goods available for trade is fixed, then there is no subgame-perfect equilibrium in which an agent obtains a positive expected discounted payoff.

We then relax assumption (a) by considering the case where the transaction cost follows a geometric Brownian motion. In this case, we derive a closed-form solution for the unique maximal symmetric subgame-perfect equilibrium, in which the agents obtain positive expected discounted payoffs by exchanging the goods through a sequence of transfers of decreasing size. Section 3.5 presents a number of comparative statics for the solution. In section 3.6, we prove an efficiency result stating that the first-best expected discounted payoffs can be approximated as the agents become infinitely patient, provided that the drift of the cost process is not excessively high relative to its volatility.

Section 3.7 examines the robustness of our results. First, in subsection 3.7.1, we relax assumption (b) by considering a model in which the supply of each good evolves according to a geometric Brownian motion. We show that if there is some volatility in the supply of each good, then positive gains from trade can be supported in a subgame-perfect equilibrium, even if the transaction cost is a positive constant. In subsection 3.7.2, we consider the case where the transaction cost depends not only on the term that evolves according to a geometric Brownian motion, but also on a term that is proportional to the amount of each good transferred. We show that positive transfers can be sustained in such a setting.

Finally, section 3.8 offers some concluding remarks. All proofs, except the one for the main result (Theorem 3.4.4), are provided in the appendix.

3.2 Related Literature

Our model is related to a line of literature on gradualism in contribution games and concession bargaining.[6] In these models, parties arrive at an agreement in a step-by-step fashion, and there is an efficiency loss due to delay in reaching an agreement. Likewise, cooperation between the two parties in our model is sustained through a gradual sequence of transactions over time. Nonetheless, our model differs from much of this literature in that transactions continue indefinitely in equilibrium; so that, our model cannot be solved using an iterated-dominance procedure. The key modeling difference from, for instance, Admati and Perry (1991) is that in their model, a benefit from cooperation is realized only when a joint project is completed, whereas in our model, a benefit is received every time a transaction takes place.

Kamada and Kandori (2009) consider the exchange of goods in the context of a revision game in which agents have an opportunity to revise their actions according to a Poisson process and payoffs are realized only upon reaching a given deadline. Those authors construct an equilibrium in which a nontrivial exchange occurs at the deadline and transfers become smaller and smaller as the deadline approaches. However, the logic behind gradually decreasing transfers is significantly different in our paper. In their model, an assumption on payoff functions implies that as actions become smaller, the instantaneous gain from deviation becomes infinitely negligible relative to the payoff from cooperation, whereas in our model, the transfers must become smaller because the stock of goods is fixed, and an infinite sequence of transfers is required to sustain cooperation. Moreover, since payoffs realize only once in Kamada and Kandori (2009), the tradeoff between the transfer size and the waiting time described in the introduction is not an issue in their paper.

The paper most closely related to ours is Pitchford and Snyder (2004). Those authors consider

[6] See Admati and Perry (1991); Compte and Jehiel (1995, 2003, 2004); Marx and Matthews (2000).

a holdup problem between a buyer and a seller in which no investment occurs in the equilibrium of the static game. In a dynamic version of their model, however, positive investment can be supported as part of an equilibrium in which the seller's investment and the buyer's repayment take place alternately.

There are a number of differences between the model in Pitchford and Snyder (2004) and the framework developed in this paper. Pitchford and Snyder (2004) consider a discrete-time model with no uncertainty in the cost of transacting with the other party or the gains from trade that can be realized. By contrast, this paper analyzes a continuous-time setting that accommodates volatility in both the cost of transferring goods between parties and the quantity of goods available for trade. Furthermore, as Pitchford and Snyder (2004) observe, the equilibrium of their model is not robust to the inclusion of a fixed cost for making a transaction. If the seller must incur a fixed cost for investing instead of a variable cost that decreases to zero with the size of the investment, then an equilibrium with positive investment by the seller cannot be supported in an equilibrium of their model. Our model provides two ways of resolving this issue. If there is uncertainty in either the fixed cost of making a transaction or the quantity of goods available for trade, then agents might be able to realize positive gains from trade as part of a subgame-perfect equilibrium.

In addition, the efficiency properties of our model are somewhat more nuanced than those of Pitchford and Snyder (2004). Those authors demonstrate that as discounting frictions disappear, the solution to their model converges to the efficient outcome. In our model, the fixed cost of making a transaction acts as an additional friction in the bargaining process. Consequently, in the limit as the discount rate approaches zero, the solution to our model converges to the efficient outcome if and only if the drift of the transaction cost process is sufficiently low relative to its volatility. Even if this condition is satisfied, the efficiency of the equilibrium is not obvious given the results in Pitchford and Snyder (2004), because the timing of transactions in our model is endogenous. As agents become infinitely patient, not only does the cost of the first transaction decrease but the expected waiting time until this transaction increases. What our efficiency result demonstrates is that the positive effect due to a lower discount rate dominates the negative effect due to a longer waiting time. In contrast, the model of Pitchford and Snyder (2004) predicts that a large amount of transfer occurs at the very first period irrespective of the degree of friction.

Finally, our model is related to the large literature on repeated games. In standard models of repeated games, agents play a stage game for infinitely many periods, and the payoff structure for the stage game is typically constant over time.[7] For such models, the folk theorem states that any individually rational and feasible payoff can be achieved in a subgame-perfect equilibrium of the supergame if the discount rate is sufficiently low. Such equilibria are sustained by using the threat of future punishment to enforce cooperation in the current period. Likewise, the loss of future utility from deviating provides an incentive for agents in our model to exchange goods with each other. Nonetheless, there are two important differences between our model and the standard framework. First, in the case where the total supply of each good is fixed, the present discounted value of the relationship must eventually decrease with time if each agent is transferring a positive quantity of her good. Hence, it is not immediately obvious that a non-degenerate equilibrium can be sustained in our model. Indeed, we obtain an impossibility result demonstrating that a positive expected discounted payoff cannot be supported in a subgame-perfect equilibrium of our model if the transaction cost is bounded below by a positive number and the quantity of each good available for trade is fixed. Second, the timing of moves is exogenously specified in standard models of repeated games, but in our model it is an endogenous strategic response to the realization of the cost process.

3.3 Model of Stochastic Transaction Costs

There are two agents, 1 and 2, who take actions in continuous time $t \in [0, \infty)$. The discount rate is $\rho > 0$. There are two divisible goods, 1 and 2. The allocation of the goods at time t is represented by $s_t = [(s_t^{11}, s_t^{12}), (s_t^{21}, s_t^{22})]$, where s_t^{ij} denotes the amount of good j that agent i possesses at time t. The total supply $q > 0$ of each good is taken to be constant over time; so that, $s_t^{1j} + s_t^{2j} = q > 0$ for $j = 1, 2$ and $t \in [0, \infty)$.[8] The initial endowment vector is assumed to be $s_0 = [(q, 0), (0, q)]$. That is, agent 1 is endowed with all of good 1, and agent 2 is endowed with all of good 2. This assumption is without loss of generality provided that $s_0^{11} = s_0^{22}$. In addition, there is a transaction cost C_t for transferring goods between the two parties, which changes over time

[7]For example, see Fudenberg and Maskin (1986). In addition, see footnote 3 for papers that analyze stochastic games.

[8]Section 3.7.1 presents a model in which the supply of each good can vary over time.

according to some stochastic process.[9] For ease of exposition, the initial state C_0 of the transaction cost is assumed to be sufficiently high (e.g., $C_0 > q/\rho$).

In every instant of time, each agent observes the current realization of the cost and chooses an amount to transfer to the other agent. Formally, let H_t denote the set of all histories up to time t, and let $h_t = (\{C_\tau, s_\tau\}_{\tau\in[0,t)}, C_t)$ be a generic element of this set. By convention, let h_0 be the null history consisting of a singleton set. Denote the set of all histories by $H = \bigcup_{t\in\mathbb{R}_+} H_t$. Then a strategy for agent $i = 1, 2$ is a function $\pi_i : H \to \mathbb{R}_+$ satisfying $\pi_i(h_t) \in [0, \hat{s}_t^{ii}]$ for $h_t \in H$ and $\hat{s}_t^{ii} = \lim_{\tau\to t^-} s_\tau^{ii}$, where $\hat{s}_t^{ii}$ denotes the amount of agent i's good remaining when she chooses a transfer at time t.[10] The transfer made at time t by agent $i = 1, 2$ is denoted by $x_t^i = \hat{s}_t^{ii} - s_t^{ii}$. Finally, let Π_i with generic element π_i represent the set of all possible strategies for agent i.

We typically restrict attention to symmetric strategies. Our definition of a symmetric strategy profile is provided below.

Definition A strategy profile π is **symmetric** if $\pi_1(h_t) = \pi_2(h_t)$ for any $h_t \in H$ such that $\{s_\tau^{11}\}_{\tau<t} = \{s_\tau^{22}\}_{\tau<t}$.

The preceding definition of symmetry requires that given any symmetric history of play up to time t, agent 1's transfer of good 1 at time t is equal to agent 2's transfer of good 2 at time t.

If $\{s_t\}$ is the time path of allocations induced by a given symmetric strategy profile π when the cost path is $\{C_t\}$, then let $t(k;\pi)$ denote the time when the k^{th} transfer is made.[11] If the strategy profile π and the cost path $\{C_t\}$ are such that fewer than k transactions occur in total, then define

[9] We do not allow for a continuous transfer of the good in which the transferred amount is zero at each moment but the integrated amount with respect to time is positive. To avoid this possibility, one could define an expanded strategy space in which at every moment of time, each agent chooses *In* or *Out* as well as the amount to transfer. A transaction can be made and a transaction cost is paid if and only if *In* is chosen.

[10] Note that the limit exists because s_t^{ii} is a monotonic function of time. Defining strategies in continuous time involves a number of difficulties in general, as discussed in Simon and Stinchcombe (1989). Nonetheless, agents take only a countable number of actions as a best response after any history because a discrete positive cost is incurred at every instance at which a transfer is made; so that, this definition of strategies is innocuous.

[11] We assume that for any $\tau > 0$, the number of transactions in time $[0, \tau]$ is finite with probability one. This restriction is justified for the cost processes considered in this paper. In particular, if this restriction is violated when strategy profile π is played, each agent would make an infinite number of transfers in a finite amount of time with positive probability, thereby obtaining an expected discounted payoff of $-\infty$. Clearly, such a π cannot constitute a subgame-perfect equilibrium.

$t(k;\pi) = \infty$. The present discounted payoff to each agent at time d is given by

$$U_d(\pi, \{C_t\}) = \int_d^\infty e^{-\rho\cdot(t-d)} s_t^{ij} dt \;\; - \sum_{\{k:\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} C_{t(k;\pi)} \qquad \text{for} \quad j \neq i. \tag{3.1}$$

Since there is no uncertainty regarding past play and past events, we use subgame-perfect equilibrium (henceforth "SPE") as our equilibrium concept. Let Π^* denote the nonempty set of all symmetric SPE.[12] In addition, when strategy profile π is played, let $\mathbb{E}[U_d(\pi, \{C_t\})|\{\tilde{C}_\tau\}_{\tau\in[0,d]}]$ denote the expected discounted payoff to each agent at time d given that the cost path up to time d is $\{\tilde{C}_\tau\}_{\tau\in[0,d]}$. Our definition of optimal behavior is as follows.

Definition A symmetric SPE π is **maximal** in the class of symmetric SPE if for every time d and cost path $\{\tilde{C}_\tau\}_{\tau\in[0,d]}$ up to time d, there is no $\pi' \in \Pi^*$ such that $\mathbb{E}[U_d(\pi', \{C_t\})|\{\tilde{C}_\tau\}_{\tau\in[0,d]}] > \mathbb{E}[U_d(\pi, \{C_t\})|\{\tilde{C}_\tau\}_{\tau\in[0,d]}]$.

There are several reasons for focusing on a maximal equilibrium.[13] First, such an equilibrium is salient. Second, we consider a situation in which the two parties have made an informal agreement with each other. Hence, it is reasonable to assume that agents can coordinate their play so as to induce their preferred outcome as long as the incentive constraints of each party are not violated. Third, restricting attention to such an equilibrium makes it possible to obtain a unique solution, which enables us to obtain meaningful comparative statics.

3.4 Analysis of Model

We begin with the following result that greatly simplifies the analysis as well as the exposition.

[12]The strategy profile in which each agent never makes a transfer conditional on any history is an element of Π^*; therefore, the set Π^* is nonempty.

[13]A weaker definition of a maximal symmetric equilibrium π would simply require there to be no $\pi' \in \Pi^*$ such that $\mathbb{E}[U_0(\pi', \{C_t\})] > \mathbb{E}[U_0(\pi, \{C_t\})]$. The analysis of the model would not change substantially under this alternative definition, except that there would be no restriction on events occurring with zero probability in equilibrium. In order to simplify the exposition by eliminating reference to zero-probability events, the stronger definition given above is used.

Proposition 3.4.1 *The expected discounted payoff $U_d(\pi, \{C_t\})$ given in equation (3.1) is proportional to*

$$s^{ij}_{t(k-1;\pi)} + \sum_{\{k:\ t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} x^{j}_{t(k;\pi)} - \sum_{\{k:\ t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} c_{t(k;\pi)},$$

where $i \neq j$, $c_t = \rho C_t$, and $t(k-1;\pi) = 0$ if $k = 1$.

In other words, with an appropriate rescaling of the cost process, each agent can be regarded as consuming the good from the other agent as soon as it is received.[14] Because of its tractability, we work directly with the process c_t in our analysis, but we always keep in mind the original maximization problem involving C_t.

The following theorem is an impossibility result. If the transaction cost is bounded below by a positive number and the quantity of each good available for trade is fixed, then there is no equilibrium in which an agent receives a positive expected discounted payoff.

Theorem 3.4.2 *Assume that $\{\tilde{c}_t\}$ is an arbitrary cost process defined on the probability space $(\Omega, \mathcal{F}, P)$ and that each random variable $\tilde{c}_t$ for $t \geq 0$ takes values in the state space $S \subset \mathbb{R}$ with $\inf(S) = \underline{c} > 0$. Then any SPE strategy profile satisfies $s^{ij}_t = 0$ for all t and $j \neq i$.*

Note that this result does not depend on assuming symmetric or maximal equilibrium strategies. Furthermore, the proof only relies on an induction argument. That is, for any given $\underline{c} > 0$ in the statement of the theorem, only a finite hierarchy of knowledge about rationality is needed to establish the proposition.

There are two approaches to resolve this issue and to proceed with the analysis. Section 3.7.1 examines a model with uncertainty in the quantity of each good available for trade, demonstrating that the impossibility result may not hold in such a setting. In the current section, we relax the assumption that the transaction cost is bounded below by a positive number while maintaining the assumption that there is no uncertainty in the quantity of goods available for trade. Specifically, *we assume that the cost process c_t follows a geometric Brownian motion $dc_t = \mu c_t dt + \sigma c_t dz_t$ with arbitrary drift μ and positive volatility $\sigma^2 > 0$.*

[14]Note that the additional term $s^{ij}_{t(k-1;\pi)}$ simply reflects the fact that agent i can have some amount of good j at time d. The presence of this term does not affect the analysis in this section.

We will restrict attention to SPE in grim-trigger strategies. The following result provides justification for this restriction. Note again that this proposition does not require symmetric or maximal equilibrium strategies to be played.

Proposition 3.4.3 *Given an arbitrary SPE, there exists an SPE in grim-trigger strategies that achieves the same equilibrium path of play.*

In particular, this proposition implies that given an arbitrary symmetric SPE, there exists a symmetric SPE in grim-trigger strategies that achieves the same continuation payoff after any history on the equilibrium path. However, the proposition has further content. It states that when characterizing the equilibrium *strategies* on the path of play, it is permissible to limit the analysis to grim-trigger strategies.

The basic idea behind the proof is that each agent can always obtain a continuation payoff of zero by transferring nothing. If an opponent uses a grim-trigger strategy, then zero is the maximum payoff that can be achieved after any deviation. Thus, the grim-trigger strategy is the most severe punishment available.

Hereafter, we restrict attention to maximal symmetric SPE in grim-trigger strategies. We can now state the main theorem of this paper, which provides a closed-form solution for the equilibrium strategies.

Theorem 3.4.4 *Any maximal symmetric SPE in grim-trigger strategies is characterized by a sequence $\{c_k^*, x_k^*\}_{k=1}^{\infty}$ satisfying*

$$c_k^* = \left(\frac{q}{1-\beta}\right)\left(\frac{\beta}{\beta-1}\right)^{k-\beta} \quad \textit{and} \quad x_k^* = \left(\frac{q}{1-\beta}\right)\left(\frac{\beta}{\beta-1}\right)^{k-1}$$

with

$$\beta = \tfrac{1}{2} - \tfrac{\mu}{\sigma^2} - \sqrt{\left(\tfrac{\mu}{\sigma^2} - \tfrac{1}{2}\right)^2 + 2\tfrac{\rho}{\sigma^2}},$$

such that the k^{th} transaction is made when the cost reaches c_k^* for the first time, and the amount x_k^* is transferred by each agent at this transaction.

Proof We show that any maximal symmetric SPE in grim-trigger strategies is characterized by the formula in the statement of the theorem. By Proposition 3.4.3, every maximal symmetric SPE must induce the same equilibrium path of play. Our proof relies on a sequence of nine lemmata that are presented below.

It is helpful to introduce some further notation. Given any history h_t such that $\{s^{11}_\tau\}_{\tau<t} = \{s^{22}_\tau\}_{\tau<t}$, let $W(h_t;\pi)$ denote the expected discounted payoff to each agent if both agents follow the symmetric strategy profile π from time t onwards. Letting $\tilde{\Pi}$ denote the set of symmetric SPE, we define $V(c_t, \hat{s}^{ii}_t) = \sup_{\tilde{\pi}\in\tilde{\Pi}} W(h_t;\tilde{\pi})$, where h_t is a history such that the cost at time t is c_t, the stock immediately before time t is $\hat{s}^{11}_t = \hat{s}^{22}_t$, and $\{s^{11}_\tau\}_{\tau<t} = \{s^{22}_\tau\}_{\tau<t}$.[15] In addition, define $Y(h_t;\pi)$ as follows. If $\pi_i(h_t) = 0$, then $Y(h_t;\pi) = W(h_t;\pi)$. If $\pi_i(h_t) > 0$, then $Y(h_t;\pi) = W(h_t;\pi) - [\pi_i(h_t) - c_t]$.

We begin by establishing a few properties of the value function $V(c, s)$ for the problem. The result below shows that $V(c, s)$ is decreasing in the size of the cost c and increasing in the stock of each good s.

Lemma 3.4.5 *$V(c, s)$ is decreasing in c and increasing in s.*

Proof To show that $V(c, s)$ is decreasing in c, choose any $\alpha \in (0, 1)$. Fix any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}^{ii}_t = s$. Let π be an arbitrary symmetric SPE with $W(h_t;\pi) > 0$. Consider the symmetric SPE π' defined as follows. First, the agents make no transfers until a history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = \alpha c$ and $\hat{s}'^{ii}_{t'} = s$ is reached. Second, upon reaching such a history $h'_{t'}$, the agents follow the symmetric strategy profile π'', behaving as if the time is t instead of t' and the history is $\hat{h}_t$ instead of $h'_{t'}$, where $\hat{h}_t = (\{\alpha c_\tau, s_\tau\}_{\tau\in[0,t)}, \alpha c_t)$. The SPE π'' is defined as follows. If $\bar{h}_u = (\{\bar{c}_\tau, \bar{s}_\tau\}_{\tau\in[0,u)}, \bar{c}_u)$ and $\tilde{h}_u = (\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u)$ are histories such that $(\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u) = (\{\alpha\bar{c}_\tau, \bar{s}_\tau\}_{\tau\in[0,u)}, \alpha\bar{c}_u)$, then $\pi''_i(\tilde{h}_u) = \pi_i(\bar{h}_u)$. Observe that $W(h'_{t'};\pi') > W(h_t;\pi)$. That is, given an arbitrary symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}^{ii}_t = s$, one can find a symmetric SPE π' such that $W(h'_{t'};\pi') > W(h_t;\pi)$ for any history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = \alpha c$ and $\hat{s}'^{ii}_{t'} = s$.

Noting that $V(c, s) > 0$, choose any ϵ such that $0 < \epsilon < V(c, s)$. Then one can find $\lambda > 0$, $\nu > 0$ such that for any symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}^{ii}_t = s$ satisfying $V(c, s) - W(h_t;\pi) < \epsilon$, the probability conditional on reaching history h_t of a transaction occurring between times t and ν inclusive must be greater than λ. Hence, one can find

[15] Note that the value function $V(c_t, \hat{s}^{ii}_t)$ is stationary, depending only on the cost c_t at time t and the stock $\hat{s}^{ii}_t$ immediately before time t.

$\kappa > 0$ such that for any symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}_t^{ii} = s$ satisfying $V(c,s) - W(h_t;\pi) < \epsilon$, the expected discounted sum of transaction costs incurred from time t onward conditional on reaching history h_t must be greater than $\kappa > 0$. It follows that $V(\alpha c, s) > V(c,s)$, where $\alpha \in (0,1)$.

To show that $V(c,s)$ is increasing in s, choose any $\alpha > 1$. Fix any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}_t^{ii} = s$. Let π be an arbitrary symmetric SPE with $W(h_t;\pi) > 0$. Consider the symmetric SPE π' defined as follows. First, the agents make no transfers until a history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = c$ and $\hat{s}'^{ii}_{t'} = \alpha s$ is reached. Second, upon reaching such a history $h'_{t'}$, the agents follow the symmetric strategy profile π'', behaving as if the time is t instead of t' and the history is $\hat{h}_t$ instead of $h'_{t'}$, where $\hat{h}_t = (\{c_\tau, \alpha s_\tau\}_{\tau\in[0,t)}, c_t)$. The SPE π'' is defined as follows. If $\bar{h}_u = (\{\bar{c}_\tau, \bar{s}_\tau\}_{\tau\in[0,u)}, \bar{c}_u)$ and $\tilde{h}_u = (\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u)$ are histories such that $(\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u) = (\{\bar{c}_\tau, \alpha\bar{s}_\tau\}_{\tau\in[0,u)}, \bar{c}_u)$, then $\pi''_i(\tilde{h}_u) = \alpha\pi_i(\bar{h}_u)$. Observe that $W(h'_{t'};\pi') > W(h_t;\pi)$. That is, given an arbitrary symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}_t^{ii} = s$, one can find a symmetric SPE π' such that $W(h'_{t'};\pi') > W(h_t;\pi)$ for any history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = c$ and $\hat{s}'^{ii}_{t'} = \alpha s$.

Noting that $V(c,s) > 0$, choose any ϵ such that $0 < \epsilon < V(c,s)$. Then one can find $\kappa > 0$ such that for any symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}_t^{ii} = s$ satisfying $V(c,s) - W(h_t;\pi) < \epsilon$, the expected discounted sum of transfers made from time t onward conditional on reaching history h_t must be greater than $\kappa > 0$. It follows that $V(c,\alpha s) > V(c,s)$, where $\alpha > 1$. ∎

The next result shows that $V(s,c)$ is homogenous of degree one. This property of the value function is a consequence of the assumption that the cost process follows a geometric Brownian motion.

Lemma 3.4.6 $V(s,c)$ *is homogeneous of degree one.*

Proof Choose any $\alpha > 0$. Fix any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}_t^{ii} = s$. Let π be an arbitrary symmetric SPE. Consider the symmetric SPE π' defined as follows. First, the agents make no transfers until a history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = \alpha c$ and $\hat{s}'^{ii}_{t'} = \alpha s$ is reached. Second, upon reaching such a history $h'_{t'}$, the agents follow strategy profile π'', behaving as if the time is t instead of t' and the history is $\hat{h}_t$ instead of $h'_{t'}$, where $\hat{h}_t = (\{\alpha c_\tau, \alpha s_\tau\}_{\tau\in[0,t)}, \alpha c_t)$. The symmetric SPE π'' is defined as follows. If $\bar{h}_u = (\{\bar{c}_\tau, \bar{s}_\tau\}_{\tau\in[0,u)}, \bar{c}_u)$ and $\tilde{h}_u = (\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u)$ are histories such that $(\{\tilde{c}_\tau, \tilde{s}_\tau\}_{\tau\in[0,u)}, \tilde{c}_u) = (\{\alpha\bar{c}_\tau, \alpha\bar{s}_\tau\}_{\tau\in[0,u)}, \alpha\bar{c}_u)$, then $\pi''_i(\tilde{h}_u) = \alpha\pi_i(\bar{h}_u)$. Note that $W(h'_{t'};\pi') = \alpha W(h_t;\pi)$. That is, given an arbitrary symmetric SPE π and any history $h_t = (\{c_\tau, s_\tau\}_{\tau\in[0,t)}, c_t)$ with $c_t = c$ and $\hat{s}^{ii}_t = s$, one can find a symmetric SPE π' such that $W(h'_{t'};\pi') = \alpha W(h_t;\pi)$ for any history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c_{t'})$ with $c'_{t'} = \alpha c$ and $\hat{s}'^{ii}_{t'} = \alpha s$. Hence, $V(\alpha c, \alpha s) \geq \alpha V(c,s)$. A symmetric argument can be used to show that $V(c,s) \geq \alpha^{-1}V(\alpha c, \alpha s)$. It follows that for any $\alpha > 0$, we have $V(\alpha c, \alpha s) = \alpha V(c,s)$. ∎

We now proceed to characterize the properties of a maximal symmetric SPE. The following result shows that in any maximal symmetric SPE, the cost paid at each transaction is equal to the continuation value immediately after the transaction. That is, the incentive constraint must bind at every transaction in a maximal symmetric SPE.

Lemma 3.4.7 *In any maximal symmetric SPE π, each transaction k must satisfy*

$$c_{t(k;\pi)} = V(c_{t(k;\pi)}, s^{ii}_{t(k;\pi)})$$

for every cost path $\{c_t\}$ with $t(k;\pi) < \infty$.

Proof Suppose that π is a maximal symmetric SPE for which there exists a transaction $\tilde{k}$ such that $c_{t(\tilde{k};\pi)} < V(c_{t(\tilde{k};\pi)}, s^{ii}_{t(\tilde{k};\pi)})$ for some cost path $\{\tilde{c}_t\}$ with $t(\tilde{k};\pi) < \infty$. If $Y(h_{t(\tilde{k};\pi)};\pi) < V(c_{t(\tilde{k};\pi)}, s^{ii}_{t(\tilde{k};\pi)})$ for cost path $\{\tilde{c}_t\}$, then π could not be a maximal symmetric SPE, because one could find another SPE π' such that $W(h_{t(\tilde{k};\pi)};\pi') > W(h_{t(\tilde{k};\pi)};\pi)$ for cost path $\{\tilde{c}_t\}$. In particular, let π' be defined as follows. The agents follow strategy profile π until they reach a history $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c'_{t'})$ with $Y(h'_{t'};\pi) < V(c'_{t'}, s'^{ii}_{t'})$ such that the $\tilde{k}^{th}$ transaction occurs at history $h'_{t'}$. Given such a history $h'_{t'}$, choose any history $h''_{t'} = (\{c''_\tau, s''_\tau\}_{\tau\in[0,t')}, c''_{t'})$ such that $c''_{t'} = c'_{t'}$ and $\hat{s}''^{ii}_{t'} = \hat{s}'^{ii}_{t'} - \pi(h'_{t'})$. Then there exists a symmetric SPE strategy profile π'' such that $Y(h'_{t'};\pi) < W(h''_{t'};\pi'') \leq V(c''_{t'}, \hat{s}''^{ii}_{t'}) = V[c'_{t'}, \hat{s}'^{ii}_{t'} - \pi'(h'_{t'})]$. Upon reaching such a history $h'_{t'}$, the agents transfer the amount $\pi(h'_{t'}) + \pi''(h''_{t'})$ at time t'. After time t', the agents follow strategy profile π'', behaving as if history $h''_{t'}$ instead of history $h'_{t'}$ were reached at time t'.

Therefore, it must be that $Y(h_{t(\tilde{k};\pi)};\pi) = V(c_{t(\tilde{k};\pi)}, s^{ii}_{t(\tilde{k};\pi)})$ for every cost path $\{c_t\}$ with $t(\tilde{k};\pi) < \infty$. Now, if $c_{t(\tilde{k};\pi)} < Y(h_{t(\tilde{k};\pi)};\pi)$ for cost path $\{\tilde{c}_t\}$, then π could not be a maximal symmetric SPE, because one could find another symmetric SPE $\hat{\pi}$ such that $W(h_{t(\tilde{k};\pi)};\hat{\pi}) >$

$W(h_{t(\tilde{k};\pi)}, \pi)$ for cost path $\{\tilde{c}_t\}$. In particular, let t_1, t_2 be any two nonnegative real numbers such that $t_1 < t_2$. Given a history h_{t_1} for which $\{s^{11}_\tau\}_{\tau \le t_1} = \{s^{22}_\tau\}_{\tau \le t_1}$, let $\Sigma(h_{t_1}, \{c_\tau\}_{\tau \in (t_1, t_2]}; t_1, t_2; \pi)$ be the sum of the transfers that would be made between times t_1 and t_2 inclusive if history h_{t_1} is reached at time t_1, the cost path $\{c_\tau\}_{\tau \in (t_1, t_2]}$ is realized between times t_1 and t_2, and strategy profile π is played by the agents.

Define the symmetric strategy profile $\hat{\pi}$ as follows. First, the agents make follow strategy profile π until they reach a history $h_{\hat{t}_1}$ with $c_{\hat{t}_1} < Y(h_{\hat{t}_1}; \pi)$ such that the $\tilde{k}^{th}$ transaction occurs at history $h_{\hat{t}_1}$. Second, when any such history $h_{\hat{t}_1}$ is reached, each agent transfers the amount $\pi_i(h_{\hat{t}_1}) + \delta(h_{\hat{t}_1})$, where $0 < \delta(h_{\hat{t}_1}) < Y(h_{\hat{t}_1}; \pi) - c_{\hat{t}_1}$. Third, the agents do not make a transaction at any time $\tilde{t}_2 > \hat{t}_1$ such that the cost path $\{c_\tau\}_{\tau \in (\hat{t}_1, \tilde{t}_2]}$ satisfies $\Sigma(h_{\hat{t}_1}, \{c_\tau\}_{\tau \in (\hat{t}_1, \tilde{t}_2]}; \hat{t}_1, \tilde{t}_2; \pi) < \pi_i(h_{\hat{t}_1}) + \delta(h_{\hat{t}_1})$. Fourth, each agent transfers the amount $\Sigma(h_{\hat{t}_1}, \{c_\tau\}_{\tau \in (\hat{t}_1, \hat{t}_2]}; \hat{t}_1, \hat{t}_2; \pi) - \pi_i(h_{\hat{t}_1}) - \delta(h_{\hat{t}_1})$ at the first time $\hat{t}_2 \ge \hat{t}_1$ such that the cost path $\{c_\tau\}_{\tau \in (\hat{t}_1, \hat{t}_2]}$ satisfies $\Sigma(h_{\hat{t}_1}, \{c_\tau\}_{\tau \in (\hat{t}_1, \hat{t}_2]}; \hat{t}_1, \hat{t}_2; \pi) \ge \pi_i(h_{\hat{t}_1}) + \delta(h_{\hat{t}_1})$. Fifth, the agents follow strategy profile π at any time $\bar{t}_2 > \hat{t}_2$, behaving as if strategy profile π had been played from history $h_{\hat{t}_1}$ onwards.

Because $\hat{\pi}$ would be a symmetric SPE such that $W(h_{t(\tilde{k};\pi)}; \hat{\pi}) > W(h_{t(\tilde{k};\pi)}, \pi)$ for cost path $\{\tilde{c}_t\}$, it could not be that π is a maximal symmetric SPE if $c_{t(\tilde{k};\pi)} < Y(h_{t(\tilde{k};\pi)}; \pi)$ for cost path $\{\tilde{c}_t\}$. Hence, in any maximal symmetric SPE π, it must be that each transaction k satisfies $c_{t(k;\pi)} = Y(h_{t(k;\pi)}; \pi) = V(c_{t(k;\pi)}, s^{ii}_{t(k;\pi)})$ every cost path $\{c_t\}$ with $t(k;\pi) < \infty$. ∎

The basic idea behind the preceding result is as follows. Suppose that the cost incurred at some transaction $\tilde{k}$ is always less than the continuation value immediately after transaction $\tilde{k}$. Then it might be possible to lower the transfer made at transaction $\tilde{k} + 1$ by some amount and to raise the transfer made at transaction $\tilde{k}$ by the same amount without violating any incentive constraints. By doing so, a future payoff would be experienced at an earlier point in time, thereby increasing the expected discounted value of the relationship. Thus, it could not be optimal for the cost incurred at transaction $\tilde{k}$ to be less than the continuation value immediately after transaction $\tilde{k}$. The actual proof is somewhat more complicated because the cost at which a transaction occurs and the amount transferred at each transaction can be stochastic.

The result below identifies a proportionality condition that any maximal symmetric SPE must satisfy. Specifically, the ratio of the cost paid at a given transaction to the amount of each good remaining after that transaction must be a constant.

Lemma 3.4.8 *In any maximal symmetric SPE π, each transaction k must satisfy*

$$s^{ii}_{t(k+1;\pi)}/s^{ii}_{t(k;\pi)} = c_{t(k+1;\pi)}/c_{t(k;\pi)}$$

for every cost path $\{c_t\}$ with $t(k+1;\pi) < \infty$.

Proof Let π be a maximal symmetric SPE. Suppose that $\tilde{k}$ is such that $(c_{t(\tilde{k};\pi)}/c_{t(\tilde{k}+1;\pi)})s^{ii}_{t(\tilde{k}+1;\pi)} > s^{ii}_{t(\tilde{k};\pi)}$ for some cost path $\{\tilde{c}_t\}$ with $t(\tilde{k}+1;\pi) < \infty$. From Lemmata 3.4.6 and 3.4.7, we have $c_{t(\tilde{k}+1;\pi)} = V(c_{t(\tilde{k}+1;\pi)}, s^{ii}_{t(\tilde{k}+1;\pi)}) = (c_{t(\tilde{k}+1;\pi)}/c_{t(\tilde{k};\pi)})V[c_{t(\tilde{k};\pi)}, (c_{t(\tilde{k};\pi)}/c_{t(\tilde{k}+1;\pi)})s^{ii}_{t(\tilde{k}+1;\pi)}]$ for cost path $\{\tilde{c}_t\}$. It follows from Lemma 3.4.5 that $V[c_{t(\tilde{k};\pi)}, (c_{t(\tilde{k};\pi)}/c_{t(\tilde{k}+1;\pi)})s^{ii}_{t(\tilde{k}+1;\pi)}] > V(c_{t(\tilde{k};\pi)}, s^{ii}_{t(\tilde{k};\pi)})$ for cost path $\{\tilde{c}_t\}$ if $(c_{t(\tilde{k};\pi)}/c_{t(\tilde{k}+1;\pi)})s^{ii}_{t(\tilde{k}+1;\pi)} > s^{ii}_{t(\tilde{k};\pi)}$ for cost path $\{\tilde{c}_t\}$. Hence, we have $c_{t(\tilde{k};\pi)} > V(c_{t(\tilde{k};\pi)}, s^{ii}_{t(\tilde{k};\pi)})$ for cost path $\{\tilde{c}_t\}$, which contradicts Lemma 3.4.7. Thus, each transaction k must satisfy $(c_{t(k;\pi)}/c_{t(k+1;\pi)})s^{ii}_{t(k+1;\pi)} \leq s^{ii}_{t(k;\pi)}$ for every cost path $\{c_t\}$ with $t(k+1;\pi) < \infty$. By a symmetric argument, each transaction k must satisfy $(c_{t(k+1;\pi)}/c_{t(k;\pi)})s^{ii}_{t(k;\pi)} \leq s^{ii}_{t(k+1;\pi)}$ for every cost path $\{c_t\}$ with $t(k+1;\pi) < \infty$. Hence, it must be that $s^{ii}_{t(k+1;\pi)}/s^{ii}_{t(k;\pi)} = c_{t(k+1;\pi)}/c_{t(k;\pi)}$ for every cost path $\{c_t\}$ with $t(k+1;\pi) < \infty$. ∎

Because $s_{t(k;\pi)} > s_{t(k+1;\pi)}$ by definition, Lemma 3.4.8 implies that $c_{t(k;\pi)} > c_{t(k+1;\pi)}$ for every cost path $\{c_t\}$ with $t(k+1;\pi) < \infty$ in a maximal symmetric SPE π. That is, the sequence of transaction costs paid in any maximal symmetric SPE must be decreasing.

The following result shows that in order to solve for a maximal symmetric SPE, attention can be restricted to strategy profiles such that the next transaction occurs when the cost reaches some fixed level. In particular, given a maximal symmetric SPE in which there is more than one possible cost level at which the next transaction can occur, one can find another maximal symmetric SPE in which there is only one possible cost level at which the next transaction can occur. Moreover, this cost level can be any one of the cost levels at which a transaction can occur in the original SPE.

Lemma 3.4.9 *Fix any transaction k. Let π be a maximal symmetric SPE for which there exists a cost path $\{c^a_t\}$ such that transaction k occurs at cost level c^*. Let t^* denote the time when transaction $k-1$ occurs if the cost path is $\{c^a_t\}$ and strategy profile π is played, where $t^* = 0$ if $k = 1$. Then one can construct a maximal symmetric SPE π' in which the following hold for every cost path $\{c^b_t\}$ with $t(k;\pi') < \infty$ such that $\{c^b_\tau\}_{\tau\in[0,t^*]} = \{c^a_\tau\}_{\tau\in[0,t^*]}$:*

1. $h^b_{t^*} = h^a_{t^*}$,

2. $\pi'_i(h^b_{t^*}) = \pi_i(h^a_{t^*})$,

3. $c^b_{t(k;\pi')} = c^*$,

where $h^a_{t^*}$ is the history up to time t^* if the cost path is $\{c^a_t\}$ and strategy profile π is played, and $h^b_{t^*}$ is the history up to time t^* if the cost path is $\{c^b_t\}$ and strategy profile π' is played.

Proof Fixing any transaction k, let π be a maximal symmetric SPE for which there exists a cost path $\{c^a_t\}$ such that transaction k occurs at cost level c^*. If $c_{t(k,\pi)} = c^{**}$ for every cost path $\{c^b_t\}$ with $t(k;\pi) < \infty$ such that $\{c^b_\tau\}_{\tau\in[0,t^*]} = \{c^a_\tau\}_{\tau\in[0,t^*]}$, then the proof is complete, because π' in the statement of the lemma can be defined to be the same as π. Therefore, assume that there exists a cost path $\{c^d_t\}$ with $\{c^d_\tau\}_{\tau\in[0,t^*]} = \{c^a_\tau\}_{\tau\in[0,t^*]}$ such that $c^d_{t(k,\pi)} = c^{**}$ with $c^{**} \neq c^*$.

For every such cost path $\{c^d_t\}$ with $c^{**} > c^*$, it must be that $W(h''_{t''};\pi) = W(h'_{t'};\pi)$, where $h'_{t'}$ is any history up to a time t' such that $c^a_{t'} = c^{**}$ and $t(k-1,\pi) < t' < t(k,\pi)$ when the cost path is $\{c^a_t\}$ and strategy profile π is played, and $h''_{t''}$ is any history up to a time t'' such that $c^d_{t''} = c^{**}$ and $t(k-1,\pi) < t'' < t(k,\pi)$ when the cost path is $\{c^d_t\}$ and strategy profile π is played. For every such cost path $\{c^d_t\}$ with $c^{**} < c^*$, it must be that $W(h''_{t''};\pi) = W(h'_{t'};\pi)$, where $h'_{t'}$ is any history up to a time t' such that $c^a_{t'} = c^*$ and $t(k-1,\pi) < t' < t(k,\pi)$ when the cost path is $\{c^a_t\}$ and strategy profile π is played, and $h''_{t''}$ is any history up to a time t'' such that $c^d_{t''} = c^*$ and $t(k-1,\pi) < t'' < t(k,\pi)$ when the cost path is $\{c^d_t\}$ and strategy profile π is played. These claims follow from two facts. First, a maximal symmetric SPE π must satisfy $W(h_t;\pi) \geq W(h_t;\hat{\pi})$ for every history h_t on the equilibrium path and every symmetric SPE $\hat{\pi}$. Second, if $h'_{t'} = (\{c'_\tau, s'_\tau\}_{\tau\in[0,t')}, c'_{t'})$ and $h''_{t''} = (\{c''_\tau, s''_\tau\}_{\tau\in[0,t'')}, c''_{t''})$ are two histories such that $c'_{t'} = c''_{t''}$ and $\hat{s}'_{t'} = \hat{s}''_{t''}$, then any path of play that can be supported as an SPE at history $h'_{t'}$ can be supported as an SPE at history $h''_{t''}$.

Hence, the maximal symmetric SPE π' in the statement of the lemma can be defined as follows. The agents play strategy profile π up to and including the time $\tilde{t}$ when the $(k-1)^{th}$ transaction occurs. If the cost path up to time $\tilde{t}$ is different from $\{c^a_\tau\}_{\tau\in[0,t^*]}$, then the agents continue to play strategy profile π after time $\tilde{t}$. If the cost path up to time $\tilde{t}$ is the same as $\{c^a_\tau\}_{\tau\in[0,t^*]}$, then the agents do not make any transfers until the first time $\hat{t}$ that the cost reaches c^*. In addition, letting $\bar{t}$ denote the time when transaction k occurs if the cost path is $\{c^a_t\}$ and strategy profile π is played, the agents follow strategy profile π from time $\hat{t}$ onwards, behaving at time $\hat{t}$ as if the time were $\bar{t}$, the cost path were $\{c^a_\tau\}_{\tau\in[0,\bar{t}]}$, and strategy profile π was always played. ∎

In addition, as the next result demonstrates, attention can further be restricted to strategy profiles such that the next transaction occurs at the first time that the cost reaches some fixed level. That is, in any maximal symmetric equilibrium such that the next transaction occurs when the cost reaches some fixed level, the next transaction must occur at the first time that the cost reaches this level.

Lemma 3.4.10 *Fix any transaction k. Let π be a maximal symmetric SPE. Fix any cost path $\{c^a_t\}$ with $t(k-1;\pi)<\infty$. Let t^* denote the time when transaction $k-1$ occurs if the cost path is $\{c^a_t\}$ and strategy profile π is played, where $t^*=0$ if $k=1$. Suppose that $c^b_{t(k,\pi)}=c^*$ for every cost path $\{c^b_t\}$ with $t(k;\pi)<\infty$ such that $\{c^b_\tau\}_{\tau\in[0,t^*]}=\{c^a_\tau\}_{\tau\in[0,t^*]}$. Then there is no cost path $\{c^d_t\}$ with $\{c^d_\tau\}_{\tau\in[0,t^*]}=\{c^a_\tau\}_{\tau\in[0,t^*]}$ for which there exists a time $\hat{t}$ such that $\hat{t}<t(k,\pi)$ and $c^d_{\hat{t}}=c^*$.*

Proof Suppose to the contrary that $\{c^d_t\}$ is a cost path with $\{c^d_\tau\}_{\tau\in[0,t^*]}=\{c^a_\tau\}_{\tau\in[0,t^*]}$ for which such a time $\hat{t}$ exists. Let $\hat{h}_{\hat{t}}$ denote the history at time $\hat{t}$ if the cost path is $\{c^d_t\}$ and strategy profile π is played. Observe that if strategy profile π is played, any possible next transaction upon reaching history $\hat{h}_{\hat{t}}$ must occur at a future history $h'_{t'}$ in which the size of the transaction cost and the remaining amount of each good are the same as at history $\hat{h}_{\hat{t}}$. Moreover, there must exist some such future history $h'_{t'}$ in which $W(h'_{t'};\pi)>W(\hat{h}_{\hat{t}};\pi)$, because $W(h_t;\pi)>0$ for every history h_t on the equilibrium path in a maximal symmetric SPE.

Now, consider the symmetric SPE π' defined as follows. The agents play strategy profile π up to time $\hat{t}$. If the cost path up to time $\hat{t}$ is different from $\{c^d_\tau\}_{\tau\in[0,\hat{t}]}$, then the agents continue to play strategy profile π from time $\hat{t}$ onwards. If the cost path up to time $\hat{t}$ is the same as $\{c^d_\tau\}_{\tau\in[0,\hat{t}]}$, then the agents follow strategy profile π from time $\hat{t}$ onwards but behave at time $\hat{t}$ as if the time is t' and the history is $h'_{t'}$.

Note that $W(\hat{h}_{\hat{t}};\pi')>W(\hat{h}_{\hat{t}};\pi)$, contradicting the fact that π is a maximal symmetric SPE. Hence, no such cost path $\{c^d_t\}$ can exist. ∎

Consider any maximal symmetric SPE π. Choose any nonnegative integer k. If $k=1$, let $h'_{t'}$ be any history before the first transaction. If $k\geq 2$, let $h'_{t'}$ be any history after transaction $k-1$ but before transaction k. For any nonnegative integer m, let $\tilde{x}_m$ be the amount of each good transferred at transaction m, $\tilde{c}_m$ be the cost at which transaction m occurs, and $\tilde{s}_m$ be the amount of each good remaining after transaction m. If $m = 0$, then simply define $\tilde{x}_0 = 0$, $\tilde{c}_0 = c_0$, and $\tilde{s}_0 = s_0^{ii}$.

From Lemmata 3.4.9 and 3.4.10, it can be assumed that history $h'_{t'}$ is such that transaction k occurs at the first time that the cost reaches $\tilde{c}_k$. Furthermore, Lemma 3.4.7 shows that the cost paid $\tilde{c}_k$ at transaction k must be equal to the continuation value $V(\tilde{c}_k, \tilde{s}_k)$ at transaction k; so that, it is only necessary to consider the instantaneous payoff $\tilde{x}_k$ at transaction k when calculating $W(h'_{t'}, \pi)$. Finally, note that $W(h'_{t'}, \pi) = V(c, \tilde{s}_{k-1})$, where c with $c \geq \tilde{c}_k$ is the size of the cost at history $h'_{t'}$.

Hence, $W(h'_{t'}, \pi)$ is the same as the value of an asset that pays an amount $\tilde{x}_k$ at the first time that the cost reaches $\tilde{c}_k$ when the size of the cost is currently c. The Bellman equation for this asset-pricing problem is given by the following for $c \geq \tilde{c}_k$:

$$\rho V(c, \tilde{s}_{k-1})dt = \mathbb{E}(dV) \tag{3.2}$$

subject to the boundary condition $V(\tilde{c}_k, \tilde{s}_{k-1}) = \tilde{x}_k$. The result below reports the solution for the value function $V(c, \tilde{s}_{k-1})$.

Lemma 3.4.11 *The value function is given by $V(c, \tilde{s}_{k-1}) = \tilde{x}_k(c/\tilde{c}_k)^\beta$ for any $c \geq \tilde{c}_k$, where β is defined in the statement of Theorem 3.4.4.*

Proof A straightforward application of Ito's lemma to equation (3.2) yields

$$\rho V(c, \tilde{s}_{k-1}) = \mu c \frac{\partial V(c, \tilde{s}_{k-1})}{\partial c} + \frac{1}{2}\sigma^2 c^2 \frac{\partial^2 V(c, \tilde{s}_{k-1})}{\partial c^2},$$

which provides a second-order linear differential equation for $V(c, \tilde{s}_{k-1})$. Seeking a solution of the form $g(c; \tilde{c}_k, \tilde{x}_k) = B(\tilde{c}_k, \tilde{x}_k)c^{\tilde{\beta}}$, the following quadratic equation is obtained by substituting the functional form into the differential equation

$$\tfrac{1}{2}\sigma^2 \tilde{\beta}(\tilde{\beta} - 1) + \mu\tilde{\beta} - \rho = 0,$$

whose solution is given by

$$\tilde{\beta} = \tfrac{1}{2} - \tfrac{\mu}{\sigma^2} \pm \sqrt{\left(\tfrac{\mu}{\sigma^2} - \tfrac{1}{2}\right)^2 + 2\tfrac{\rho}{\sigma^2}}.$$

Letting $\tilde{\beta}^{+}$ and $\tilde{\beta}^{-}$ respectively denote the positive and negative roots of the quadratic, the general solution to the differential equation is given by $V(c, \tilde{s}_{k-1}) = B^{+}(\tilde{c}_k, \tilde{x}_k)c^{\tilde{\beta}^{+}} + B^{-}(\tilde{c}_k, \tilde{x}_k)c^{\tilde{\beta}^{-}}$. It must be the case that $B^{+}(\tilde{c}_k, \tilde{x}_k) = 0$, because $V(c, \tilde{s}_{k-1})$ would otherwise become unboundedly large in absolute value as c goes to ∞. Moreover, the boundary condition $V(c, \tilde{s}_{k-1}) = \tilde{x}_k$ yields $B^{-}(\tilde{c}_k, \tilde{x}_k) = \tilde{x}_k/\tilde{c}_k^{\tilde{\beta}^{-}}$. Hence, the solution to the Bellman equation in expression (3.2) is

$$V(c, \tilde{s}_{k-1}) = \tilde{x}_k(c/\tilde{c}_k)^{\tilde{\beta}^{-}}.$$

Setting $\beta = \tilde{\beta}^{-}$, the proof is complete. ∎

Define $\tilde{r}_k = \tilde{c}_k/\tilde{s}_k$ if $k \geq 1$ and $\tilde{y}_k = \tilde{s}_k/\tilde{s}_{k-1}$ if $k \geq 2$. Recall from Lemma 3.4.8 that if $k \geq 2$, then $\tilde{s}_k/\tilde{s}_{k-1} = \tilde{c}_k/\tilde{c}_{k-1}$ or, equivalently, $\tilde{r}_{k-1} = \tilde{r}_k$. The next result solves for the values of $\tilde{r}_k$ with $k \geq 1$ and $\tilde{y}_k$ with $k \geq 2$ in a maximal symmetric equilibrium π.

Lemma 3.4.12 *In any maximal symmetric SPE, the following must hold:*

1. $\tilde{r}_k = [-1/(\beta - 1)] \cdot [(\beta - 1)/\beta]^{\beta}$ *for* $k \geq 1$,
2. $\tilde{y}_k = \beta/(\beta - 1)$ *for* $k \geq 2$,
3. $\tilde{c}_1 = \arg\max_{\tilde{c}} \, (s_0^{ii} - \tilde{c}/\tilde{r}_1)(c_0/\tilde{c})^{\beta}$,

where β *is defined in the statement of Theorem 3.4.4.*

Proof From Lemma 3.4.11, we have $V(\tilde{c}_{k-1}, \tilde{s}_{k-1}) = (\tilde{s}_{k-1} - \tilde{s}_k)(\tilde{c}_{k-1}/\tilde{c}_k)^{\beta}$ for $k \geq 2$. Substituting $\tilde{c}_k/\tilde{c}_{k-1} = \tilde{s}_k/\tilde{s}_{k-1}$ from Lemma 3.4.8, we can rewrite this as $V(\tilde{c}_{k-1}, \tilde{s}_{k-1}) = \tilde{s}_{k-1}(1 - \tilde{y}_k)(1/\tilde{y}_k)^{\beta}$. Since $V(\tilde{c}_{k-1}, \tilde{s}_{k-1})$ is strictly concave in $\tilde{y}_k$, $V(\tilde{c}_{k-1}, \tilde{s}_{k-1})$ is maximized at a unique value of $\tilde{y}_k$, which is $\tilde{y}_k = \beta/(\beta - 1)$. Hence, the largest possible value of $V(\tilde{c}_{k-1}, \tilde{s}_{k-1})$ in a maximal symmetric SPE occurs when $\tilde{y}_k = \beta/(\beta - 1)$.

In addition, using Lemma 3.4.7, $\tilde{r}_{k-1}$ can be expressed as follows for $k \geq 2$ if $\tilde{y}_k = \beta/(\beta - 1)$:

$$\tilde{r}_{k-1} = \tilde{c}_{k-1}/\tilde{s}_{k-1} = V(\tilde{c}_{k-1}, \tilde{s}_{k-1})/\tilde{s}_{k-1} = (1 - \tilde{y}_k)(1/\tilde{y}_k)^{\beta} = [-1/(\beta - 1)] \cdot [(\beta - 1)/\beta]^{\beta}.$$

From Lemma 3.4.11, the incentive constraint for each transaction can be written as $\tilde{c}_{k-1} \leq \tilde{x}_k(\tilde{c}_{k-1}/\tilde{c}_k)^{\beta}$ for $k \geq 2$, which is equivalent to $\tilde{r}_{k-1} \leq (1 - \tilde{y}_k)(1/\tilde{y}_k)^{\beta}$. This inequality binds when $\tilde{r}_{k-1}$ and $\tilde{y}_k$ are as specified above. Hence, these choices of $\tilde{r}_{k-1}$ and $\tilde{y}_k$ can be supported as part of a symmetric SPE.

It follows that any maximal symmetric SPE must be characterized by the choices of $\tilde{r}_{k-1}$ and $\tilde{y}_k$ in the statement of the lemma. In addition, $\tilde{c}_1$ must be chosen to maximize $V(\tilde{c}_0, \tilde{s}_0)$, which can be written using Lemma 3.4.11 as $\tilde{x}_1(\tilde{c}_0/\tilde{c}_1)^\beta$ or, equivalently, $(s_0^{ii} - \tilde{c}_1/\tilde{r}_1)(\tilde{c}_0/\tilde{c}_1)^\beta$. ∎

Finally, the result below solves for the sequence of costs paid and amounts transferred in a maximal symmetric SPE.

Lemma 3.4.13 *In any maximal symmetric SPE, $\tilde{c}_k$ and $\tilde{x}_k$ are given by c_k^* and x_k^* in the statement of Theorem 3.4.4.*

Proof From Lemma 3.4.12, we have $\tilde{c}_1 = \arg\max_{\tilde{c}} (s_0^{ii} - \tilde{c}/\tilde{r}_1)(c_0/\tilde{c})^\beta$. The unique solution to this maximization problem is $\tilde{c}_1 = s_0^{ii}\tilde{r}_1[\beta/(\beta-1)]$. Using Lemma 3.4.12 to substitute for $\tilde{r}_1$, we obtain an explicit formula for $\tilde{c}_1$.

Recalling that $\tilde{y}_k = \tilde{c}_k/\tilde{c}_{k-1}$, we have $\tilde{c}_k = \tilde{c}_1 \prod_{m=2}^{k} \tilde{y}^m$ for $k \geq 2$. Hence, $\tilde{c}_k$ can be calculated for $k \geq 2$ using the formula for $\tilde{c}_1$ as well as Lemma 3.4.12. Recalling that $\tilde{r}_k = \tilde{c}_k/\tilde{s}_k$, we have $\tilde{s}_k = \tilde{c}_k/\tilde{r}_k$ for $k \geq 1$. Hence, $\tilde{s}_k$ can be calculated for $k \geq 1$ using the formula for $\tilde{c}_k$ as well as Lemma 3.4.12. Finally, $\tilde{x}_k$ can be calculated for $k \geq 1$ using the formula $\tilde{x}_k = \tilde{s}_{k-1} - \tilde{s}_k$.

Performing these calculations yields the solution in the statement of Theorem 3.4.4. ∎

3.5 Comparative Statics

The closed-form solution for the maximal symmetric SPE in Theorem 3.4.4 enables us to obtain a number of comparative statics. These results are enumerated in the corollaries below.

The result below states that the cost paid must decrease with each transaction.

Corollary 3.5.1 *c_k^* is decreasing in k.*

This finding is obvious given Theorem 3.4.4. To understand the intuition for this result, observe from Lemma 3.4.7 that the continuation value after each transaction must be equal to the transaction cost incurred. In addition, recall from Lemma 3.4.5 that the value function for the problem is decreasing in the current size of the transaction cost and increasing in the remaining stock of each good. Hence, the sequence of transaction costs incurred must be decreasing, because the remaining stock of each good falls with each transaction.

The next result states that both the amount of each good transferred and the cost paid at each transaction converge to zero in the limit as the number of past transactions goes to infinity.

Corollary 3.5.2 $\lim_{k\to\infty} x_k^* = \lim_{k\to\infty} c_k^* = 0$.

This result follows directly from Theorem 3.4.4. Intuitively, if the amount transferred did not converge to zero, then the stock of each good would be completely exhausted with positive probability. Consequently, if the cost incurred did not converge to zero while the amount transferred did, then the incentive constraint would be violated for some transaction, because the expected discounted value of future transfers would be less than the size of the transaction cost to be paid.

The result below shows that all of each good will be exchanged with probability one in the limit as time goes to infinity, provided that the drift of the cost process is not excessively high relative to the volatility.

Corollary 3.5.3 *If $\mu < \frac{\sigma^2}{2}$, then* $\operatorname{plim}_{t\to\infty} s_t^{12} = \operatorname{plim}_{t\to\infty} s_t^{21} = s_0^{ii}$.

The proof simply involves calculating the sums of infinite series and is therefore omitted. This conclusion is in part a consequence of the maximality assumption. In particular, given an SPE in which the amount of each good transferred is at most $\tilde{s} < s_0^{ii}$, one can construct a Pareto superior SPE in which the agents transfer the additional amount $s_0^{ii} - \tilde{s}$ at the first transaction.

The next corollary describes how the size of each transfer changes with the drift μ and volatility σ^2 of the cost process.

Corollary 3.5.4 *If $k-1 > |\beta|$, then x_k^* is increasing in μ and decreasing in σ^2. If $k-1 < |\beta|$, then x_k^* is decreasing in μ and increasing in σ^2.*

This is an intuitive result. If the drift μ of the cost process decreases, then the cost is more likely to fall enough in the near future for the agents to make another transaction. The greater proximity of a future transaction raises the continuation value of the relationship and relaxes the incentive constraints for the problem; so that, agents can make larger transfers at early stages and smaller transfers at later stages. If the volatility σ^2 of the cost process increases, then both extremely high and low realizations of the cost process become more likely. Because the solution has a cutoff form, the favorable impact of low cost realizations dominates the adverse impact of high cost realizations. This option-value argument suggests that a high volatility σ^2 has a similar effect on the solution as a low drift μ.

The corollary below characterizes the impact of the drift μ and volatility σ^2 of the cost process on the sequence of transaction costs paid.

Corollary 3.5.5 *For all k, c_k^* is decreasing in μ and increasing in σ^2.*

Intuitively, if the drift μ decreases, then the continuation value of the relationship rises; so that, higher cost payments can be elicited from each agent without violating the incentive constraints. As has already been noted, a high volatility σ^2 has an effect similar to a low drift μ.

The following corollary examines how the transfers made and costs incurred behave in the limits as the discount rate goes to zero and to infinity.

Corollary 3.5.6 $\lim_{\rho\to 0} x_1^* = s_0^{ii}$, $\lim_{\rho\to 0} x_k^* = 0$ *for* $k \geq 2$, *and* $\lim_{\rho\to 0} c_k^* = 0$ *for all* k. *In addition,* $\lim_{\rho\to\infty} x_k^* = \lim_{\rho\to\infty} c_k^* = 0$ *for all* k.

These results follow immediately from Theorem 3.4.4. On the one hand, as the agents become infinitely patient, it is optimal for them to wait for an extremely low cost realization before making a transfer. Furthermore, a large initial transfer can be supported in equilibrium, because the incentive constraints are weak for low values of the transaction cost and so a high continuation value is not needed to prevent deviation. On the other hand, as the agents become infinitely impatient, it is impossible to induce them to incur a large transaction cost, because the continuation value from the relationship is low and so the incentive to deviate is high. The size of each transfer must also be small, in order to ensure that the continuation value is sufficiently high to sustain cooperation.

Observe that both extremely high and low values of the discount rate lead to a small cost cutoff c_k^* for each k. Thus, the expected waiting time to reach the allocation $[(q-K, K), (K, q-K)]$ for any fixed K with $0 < K < q$ is high in these extreme cases. This observation suggests that the expected waiting time is non-monotonic in the discount rate. Letting $c^*(K)$ denote the value of the transaction cost at which the amount transferred by each agent first exceeds K, Figures 3.1 to 3.4 plot $c^*(K)$ against β for $q = 50$ and $K \in \{10, 20, 30, 40\}$.

Note that β is decreasing in μ and ρ as well as increasing in σ^2. Moreover, the expected waiting time rises as $c^*(K)$ falls, becoming infinite in the limit as $c^*(K)$ goes to zero. The discontinuities in the graphs arise because the number of transactions needed for the amount transferred by each agent to first exceed K changes at such points. Specifically, the right-most curve describes the case where a single transaction is needed for the amount transferred by each agent to first exceed K. This situation occurs when μ is low, σ^2 is high, and ρ is low. In general, the k^{th} curve from the right corresponds to the case where k transactions are required for the amount transferred by each agent to first exceed K.

Furthermore, several of the comparative statistics in this section can be interpreted in terms of the applications described in the introduction. For instance, consider the example of prisoner exchange between two hostile parties, where the transaction cost represents the level of tensions between the opponents. The solution to the model suggests a prisoner exchange protocol in which the prevailing level of tensions and the quantity of prisoners traded are smaller at the next transfer than at the current transfer. In addition, if the level of tensions decreases more quickly or becomes more volatile, then each transfer of prisoners would occur at a higher level of current tensions. These changes in the environment would also result in larger transfers earlier in the relationship and smaller transfers later in the relationship. An interesting insight to arise from the analysis is that greater volatility in the level of tensions often has the same effect on the transaction scheme as a more rapid decrease in the level of tensions.

3.6 Welfare Properties

Since the expected waiting time until the occurrence of a transaction becomes infinite as the discount rate approaches zero, it is not immediate that the first-best outcome can be approximated as agents become infinitely patient. The next theorem confirms that the second-best payoff converges to the first-best payoff in the limit as the discount rate approaches zero, provided that the drift of the cost process is not excessively high relative to the volatility.

Theorem 3.6.1 *In the limit as ρ approaches zero, the expected payoff to each agent in the maximal symmetric SPE of Theorem 3.4.4 converges to the first-best outcome if and only if $\mu \leq \frac{\sigma^2}{2}$.*

On the one hand, if trading frictions are low in the sense that the drift of the cost process is sufficiently small relative to the volatility, then the first-best outcome can be approximated as the the agents become infinitely patient. On the other hand, if the drift is high and the volatility is low,

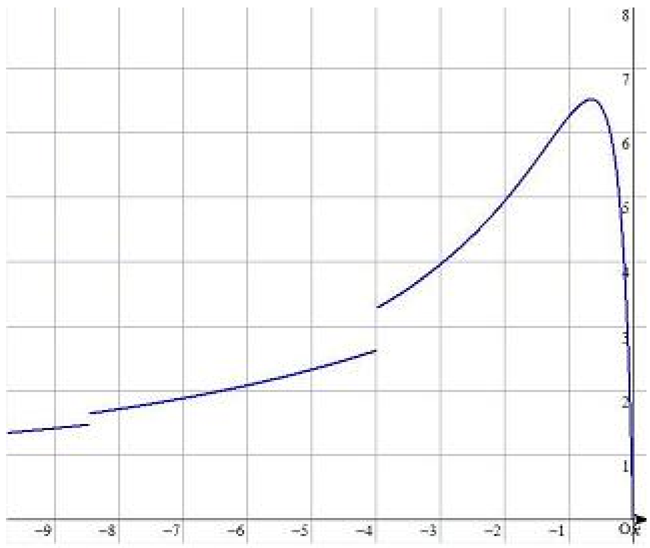

Figure 3.1: $c^*(K)$ vs. β for $(K, q) = (10, 50)$

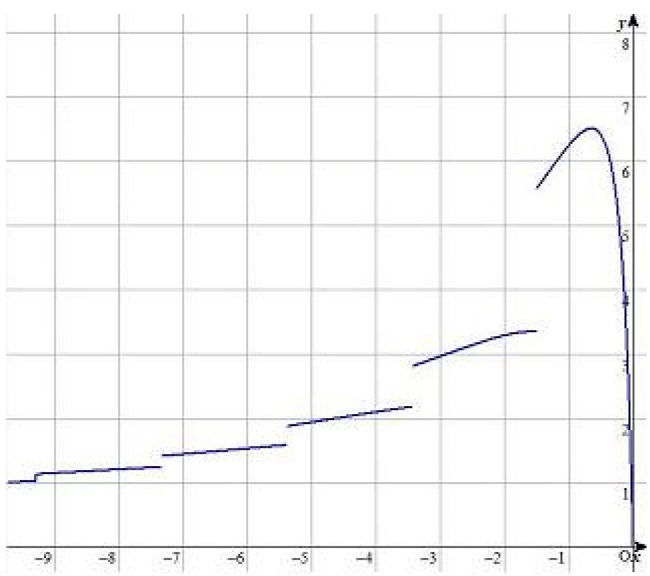

Figure 3.2: $c^*(K)$ vs. β for $(K, q) = (20, 50)$

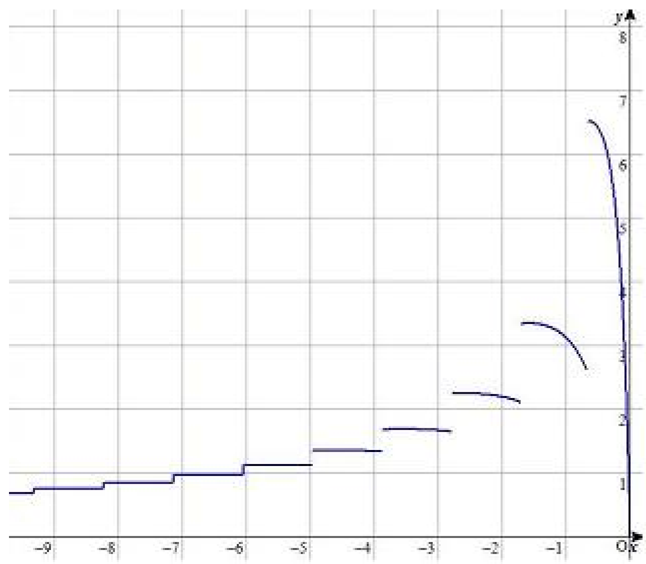

Figure 3.3: $c^*(K)$ vs. β for $(K, q) = (30, 50)$

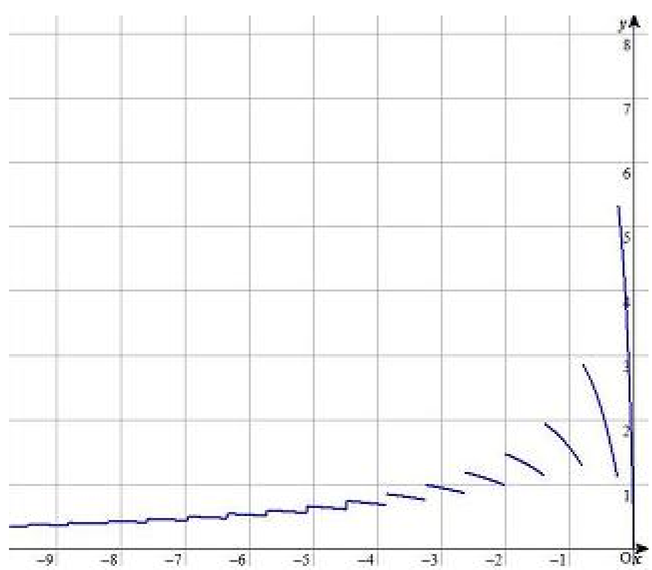

Figure 3.4: $c^*(K)$ vs. β for $(K, q) = (40, 50)$

then the trading environment tends to deteriorate over time, and there is uncertainty about whether a future transaction will take place. In this case, the second-best solution fails to converge to the first-best outcome.

The proof of preceding theorem results in the following corollary.

Corollary 3.6.2 *If $\mu > \frac{\sigma^2}{2}$, then in the limit as ρ approaches zero, the ratio of the second-best payoff to the first-best payoff is decreasing in μ and increasing in σ^2.*

Intuitively, a lower drift or a higher volatility results in a more favorable trading environment in which future transactions are more likely to occur. Consequently, the continuation value from the relationship is higher, and larger transfers can be sustained in equilibrium.

3.7 Variations of the Basic Model

This section describes two variations of the basic model. In section 3.7.1, we allow for growth or volatility in the supply of each good available for trade. We show that even if the transaction cost is bounded away from zero, positive transfers can be supported in equilibrium, provided that there is some volatility in the supply of each good. In section 3.7.2, we study a setting in which the cost incurred at a given transaction depends on the amount of each good transferred. Even in such a situation, positive transfers can be supported in equilibrium.

3.7.1 Model with Stochastic Supplies of Goods

This section presents a model with growth or volatility in the supply of each good. The setup here is the same as that described in section 3.3, except that the cost of making a transaction is fixed at some positive constant and the stock of each good is assumed to follow a geometric Brownian motion between transactions.[16] Specifically, s_t^{ii} evolves between transactions according to the equation of motion $ds_t^{ii} = \theta s_t^{ii} dt + \xi s_t^{ii} dz_t$, where z_t is a Wiener process, $\theta < \rho$ is the growth rate of each good, and $\xi^2 > 0$ is the volatility in the stock of each good. In addition, the cost of making a transaction is constant at $\chi > 0$.

[16] In addition, it is assumed as in section 3.4 that each agent consumes the good from the other agent as soon as it is received.

The following result shows that the model with a stochastic supply of each good has an equilibrium in which a transaction is made along the path of play with positive probability.

Theorem 3.7.1 *There exists a symmetric SPE of the model with stochastic supplies of goods in which the expected discounted payoff to each agent is positive at every point along the path of play.*

Three remarks are in order. First, the theorem is still valid if $\theta > \rho$. However, the problem is not well defined in this case, because there exists an SPE that generates an arbitrarily large expected discounted payoff. Second, even in the case with $\theta < 0$ where each good decays over time, a positive expected discounted payoff can be supported in equilibrium, provided that there is some volatility in the stock of each good. Although each good is decaying, there is a positive probability that the stock of each good will rise enough for another transfer to be made.

Finally, the result offers a counterpoint to Theorem 3.4.2 in the following sense. Consider a cost process with zero drift and zero volatility. If the supply of each good is fixed over time, then this cost process satisfies the conditions in Theorem 3.4.2; so that, no transactions can occur in any SPE of the model. However, if there is some volatility in the stock of each good as in the model outlined here, then the result above states that a transaction can occur in equilibrium with positive probability.

3.7.2 Model with Cost Proportional to Amount Transferred

This section examines the effect of letting the transaction cost paid depend on the amount transferred. The setup here is the same as that described in section 3.3, except that the cost paid at each transaction includes a term that is proportional to the amount transferred.[17] Specifically, if an agent transfers the amount x_t^i at time t and the fixed cost of making a transfer is c_t at time t, then the agent incurs the transaction cost $c_t + \phi \cdot x_t^i$ at time t, where $\phi \in (0, 1)$.[18]

The following result shows that as long as the component of the cost proportional to the amount transferred is not excessively large, there exists an equilibrium of the model above such that a

[17] In addition, it is assumed as in section 3.4 that each agent consumes the good from the other agent as soon as it is received.

[18] The fixed cost c_t is assumed to follow a geometric Brownian motion as in the basic model. Note that the current model would reduce to the basic model if $\phi = 0$.

transaction is made along the path of play with positive probability.

Proposition 3.7.2 *There exists $\bar{\phi} > 0$ such that if $\phi < \bar{\phi}$, then the model with a cost proportional to the amount transferred has a symmetric SPE in which the expected discounted payoff to each agent is positive at every point along the path of play.*

This result demonstrates that the implementation of positive transfers in equilibrium is robust to the introduction of a transaction cost that depends on the amount transferred. That is, it is not crucial to assume that the transaction cost does not vary with the amount transferred in order to support an equilibrium with positive gains from trade.

3.8 Conclusion

This paper studies the exchange of divisible goods between two agents facing a stochastic transaction cost. We develop a model of trade in which the first-best policy requires a single transfer of each good and the second-best policy requires a decreasing sequence of transfers of each good. We derive several comparative statics for the solution to the model, and we examine the convergence of the second-best to the first-best outcome. We also explain how the framework in this paper can be applied to real-world situations involving the gradual exchange of trade secrets, captured prisoners, or land claims.

In addition, we perform various robustness checks of the main results. We identify cases in which positive gains from trade cannot be supported in equilibrium. We also consider a model with uncertainty in the stock of each good available for trade as well as a model in which the transaction cost paid varies with the amount of each good transferred. Further extensions of the analysis might involve introducing asymmetries between the strategies followed by each agent or allowing for more general cost processes and payoff functions.

Appendix A

Appendices to Chapter 1

A.1 Proofs of Main Theoretical Results

This appendix contains the proofs of the propositions in section 1.2. The proofs of the other results discussed in the text are provided in the supplemental appendices.

A.1.1 Proof of Proposition 1.2.1

The conditional expectation of $(z_1, z_2)'$ given $(s_1, s_2)'$ is:

$$\mathbb{E}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mid \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] = \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \end{pmatrix} + \mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} - \begin{pmatrix} \mu_{s1} \\ \mu_{s2} \end{pmatrix}\right], \quad \text{(A.1)}$$

where the regression coefficient is given by:

$$\mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right]^{-1} = \left[\theta_s \sigma_s^2 \begin{pmatrix} 1 & \rho_s \\ \rho_s & 1 \end{pmatrix} + \theta_a \gamma \sigma_a^2 \begin{pmatrix} 1 & \rho_a \\ \rho_a & 1 \end{pmatrix}\right] \begin{pmatrix} \sigma_s^2 & \rho_s \sigma_s^2 \\ \rho_s \sigma_s^2 & \sigma_s^2 \end{pmatrix}^{-1}. \quad \text{(A.2)}$$

Inverting the variance matrix and rearranging terms leads to the formula in equation (1.6).

A.1.2 Proof of Proposition 1.2.2

Expressing the regression coefficient in equation (1.6) as:

$$\mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right]^{-1} = \begin{pmatrix} \theta_s + \theta_a\delta_o & \theta_a\delta_f \\ \theta_a\delta_f & \theta_s + \theta_a\delta_o \end{pmatrix}, \tag{A.3}$$

the component of $(z_1, z_2)'$ orthogonal to $(s_1, s_2)'$ is given by:

$$\begin{aligned} \begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix} &= \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} - \mathbb{E}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mid \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}\right] \\ &= \begin{pmatrix} \theta_a[a_1 - (\delta_o s_1 + \delta_f s_2)] + \omega_1 \\ \theta_a[a_2 - (\delta_f s_1 + \delta_o s_2)] + \omega_2 \end{pmatrix} - \begin{pmatrix} \theta_a[\mu_{a1} - (\delta_o\mu_{s1} + \delta_f\mu_{s2})] \\ \theta_a[\mu_{a2} - (\delta_f\mu_{s1} + \delta_o\mu_{s2})] \end{pmatrix}, \end{aligned} \tag{A.4}$$

where equations (1.5) and (A.1) are used to substitute for $(z_1, z_2)'$ and $\mathbb{E}[(z_1, z_2)'|(s_1, s_2)']$, respectively. Note that the coefficient on $(z_1, z_2)'$ in a regression on $(s_1, s_2, z_1, z_2)'$ is the same as the coefficient on $(\hat{z}_1, \hat{z}_2)'$ in a regression on $(s_1, s_2, \hat{z}_1, \hat{z}_2)'$. Therefore, consider the regression of $(a_1, a_2)'$ on $(s_1, s_2, \hat{z}_1, \hat{z}_2)'$. Because $(\hat{z}_1, \hat{z}_2)'$ is uncorrelated with $(s_1, s_2)'$, the coefficient on $(\hat{z}_1, \hat{z}_2)'$ is simply given by:

$$\begin{pmatrix} \pi_o & \pi_f \\ \pi_f & \pi_o \end{pmatrix} = \mathbb{C}\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right]^{-1}, \tag{A.5}$$

where the inverse variance matrix can be expressed as:

$$\mathbb{V}\left[\begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right]^{-1} = [\sigma_{\hat{z}}^4(1 - \rho_{\hat{z}}^2)]^{-1} \begin{pmatrix} \sigma_{\hat{z}}^2 & -\rho_{\hat{z}}\sigma_{\hat{z}}^2 \\ -\rho_{\hat{z}}\sigma_{\hat{z}}^2 & \sigma_{\hat{z}}^2 \end{pmatrix}. \tag{A.6}$$

I first show that $\mathbb{C}(a_1, \hat{z}_1) = \mathbb{C}(a_2, \hat{z}_2) > 0$. Using equations (A.4) and (1.6), this covariance is given by:

$$\mathbb{C}(a_1, \hat{z}_1) = \mathbb{C}(a_2, \hat{z}_2) = \theta_a\sigma_a^2[1 - \gamma(\delta_o + \delta_f\rho_a)] = \theta_a\sigma_a^2\left(1 - \frac{\gamma^2\sigma_a^2(1 - 2\rho_a\rho_s + \rho_a^2)}{\sigma_s^2(1 - \rho_s^2)}\right). \tag{A.7}$$

From equation (A.7), the statement $\mathbb{C}(a_1, \hat{z}_1) = \mathbb{C}(a_2, \hat{z}_2) > 0$ is equivalent to:

$$\sigma_s^2(1-\rho_s^2) - \gamma^2\sigma_a^2(1-2\rho_a\rho_s+\rho_a^2) > 0, \tag{A.8}$$

which can be expanded as:

$$k_o = (\gamma^2\sigma_a^2+\sigma_\epsilon^2)\left(1-\frac{(\gamma^2\rho_a\sigma_a^2+\rho_\epsilon\sigma_\epsilon^2)^2}{(\gamma^2\sigma_a^2+\sigma_\epsilon^2)^2}\right) - \gamma^2\sigma_a^2\left(1-2\rho_a\frac{\gamma^2\rho_a\sigma_a^2+\rho_\epsilon\sigma_\epsilon^2}{\gamma^2\sigma_a^2+\sigma_\epsilon^2}+\rho_a^2\right) > 0. \tag{A.9}$$

Note that $k_o = 0$ when $\sigma_\epsilon^2 = 0$. The derivative of k_o with respect to σ_ϵ^2 is:

$$\frac{\partial k_o}{\partial \sigma_\epsilon^2} = \frac{\sigma_\epsilon^2(1-\rho_\epsilon^2)+2\gamma^2\sigma_a^2\sigma_\epsilon^2(1-\rho_\epsilon^2)+\gamma^4\sigma_a^4(1-\rho_a^2)}{(\gamma^2\sigma_a^2+\sigma_\epsilon^2)^2} > 0. \tag{A.10}$$

It follows that $k_o > 0$ if $\sigma_\epsilon^2 > 0$ and so $\mathbb{C}(a_1, \hat{z}_1) = \mathbb{C}(a_2, \hat{z}_2) > 0$ in equation (A.7).

I next show that $\mathbb{C}(a_1, \hat{z}_2) = \mathbb{C}(a_2, \hat{z}_1) > 0$. Using equations (A.4) and (1.6), this covariance is given by:

$$\begin{aligned}\mathbb{C}(a_1, \hat{z}_2) = \mathbb{C}(a_2, \hat{z}_1) &= \theta_a\sigma_a^2[\rho_a - \gamma(\delta_o\rho_a+\delta_f)] \\ &= \theta_a\sigma_a^2\left(\rho_a - \frac{\gamma^2\sigma_a^2[2\rho_a-\rho_s(1+\rho_a^2)]}{\sigma_s^2(1-\rho_s^2)}\right).\end{aligned} \tag{A.11}$$

From equation (A.11), the statement $\mathbb{C}(a_1, \hat{z}_2) = \mathbb{C}(a_2, \hat{z}_1) > 0$ is equivalent to:

$$\rho_a\sigma_s^2(1-\rho_s^2) - \gamma^2\sigma_a^2[2\rho_a-\rho_s(1+\rho_a^2)] > 0, \tag{A.12}$$

which can be expanded as:

$$\begin{aligned}k_f =& \rho_a(\gamma^2\sigma_a^2+\sigma_\epsilon^2)\left(1-\frac{(\gamma^2\rho_a\sigma_a^2+\rho_\epsilon\sigma_\epsilon^2)^2}{(\gamma^2\sigma_a^2+\sigma_\epsilon^2)^2}\right) \\ &- \gamma^2\sigma_a^2\left(2\rho_a-(1+\rho_a^2)\frac{\gamma^2\rho_a\sigma_a^2+\rho_\epsilon\sigma_\epsilon^2}{\gamma^2\sigma_a^2+\sigma_\epsilon^2}\right) > 0.\end{aligned} \tag{A.13}$$

Note that $k_f = 0$ when $\sigma_\epsilon^2 = 0$. The derivative of k_f with respect to σ_ϵ^2 is:

$$\frac{\partial k_f}{\partial \sigma_\epsilon^2} = \frac{\gamma^4\sigma_a^4\rho_\epsilon(1-\rho_a^2)+\sigma_\epsilon^2(2\gamma^2\sigma_a^2+\sigma_\epsilon^2)\rho_a(1-\rho_\epsilon^2)}{(\gamma^2\sigma_a^2+\sigma_\epsilon^2)^2} > 0. \tag{A.14}$$

It follows that $k_f > 0$ if $\sigma_\epsilon^2 > 0$ and so $\mathbb{C}(a_1, \hat{z}_2) = \mathbb{C}(a_2, \hat{z}_1) > 0$ in equation (A.11).

I now show that $\mathbb{C}(a_1, \hat{z}_1) > \mathbb{C}(a_1, \hat{z}_2)$. From equations (A.7) and (A.11), the statement $\mathbb{C}(a_1, \hat{z}_1)$
$> \mathbb{C}(a_1, \hat{z}_2)$ is equivalent to:

$$1 - \frac{\gamma^2 \sigma_a^2 (1 - 2\rho_a \rho_a + \rho_a^2)}{\sigma_s^2 (1 - \rho_s^2)} > \rho_a - \frac{\gamma^2 \sigma_a^2 (2\rho_a - \rho_a^2 \rho_s - \rho_s)}{\sigma_s^2 (1 - \rho_s^2)}, \tag{A.15}$$

which is satisfied whenever:

$$w = \sigma_s^2 (1 - \rho_s) - \gamma^2 \sigma_a^2 (1 - \rho_a) = (\gamma^2 \sigma_a^2 + \sigma_\epsilon^2) \left(1 - \frac{\gamma^2 \rho_a \sigma_a^2 + \rho_\epsilon \sigma_\epsilon^2}{\gamma^2 \sigma_a^2 + \sigma_\epsilon^2} \right) - \gamma^2 \sigma_a^2 (1 - \rho_a) > 0. \tag{A.16}$$

Note that $w = 0$ if $\sigma_\epsilon^2 = 0$. Differentiating w with respect to σ_ϵ^2 yields:

$$\frac{\partial w}{\partial \sigma_\epsilon^2} = 1 - \rho_\epsilon > 0. \tag{A.17}$$

Hence, $w > 0$ if $\sigma_\epsilon^2 > 0$ and so $\mathbb{C}(a_1, \hat{z}_1) > \mathbb{C}(a_1, \hat{z}_2)$.

It is now straightforward to prove the three claims in proposition 1.2.2. Given the form of the inverse variance matrix in equation (A.6), it follows from $\mathbb{C}(a_1, \hat{z}_1) > \mathbb{C}(a_1, \hat{z}_2)$ that $\pi_o > \pi_f$, proving the first claim. From equations (A.5) and (A.6), the regression parameters π_o and π_f take the form:

$$\pi_o = \kappa_o - \rho_{\hat{z}} \kappa_f \quad \text{and} \quad \pi_f = \kappa_f - \rho_{\hat{z}} \kappa_o, \tag{A.18}$$

where $\kappa_o = \tau \mathbb{C}(a_1, \hat{z}_1) > 0$, $\kappa_f = \tau \mathbb{C}(a_1, \hat{z}_2) > 0$, and $\tau > 0$. Because it has been shown above that $\mathbb{C}(a_1, \hat{z}_1) > \mathbb{C}(a_1, \hat{z}_2) > 0$, one has $\kappa_o > \kappa_f > 0$. These inequalities imply that $\pi_o > 0$ in equation (A.18), proving the second claim. Finally, because $\kappa_o^2 > \kappa_f^2$ from the preceding inequalities, one has:

$$\begin{aligned} &\kappa_o^2 + \rho_{\hat{z}}^2 \kappa_f^2 > \rho_{\hat{z}}^2 \kappa_o^2 + \kappa_f^2 \Leftrightarrow \kappa_o^2 - 2\rho_{\hat{z}} \kappa_o \kappa_f + \rho_{\hat{z}}^2 \kappa_f^2 > \rho_{\hat{z}}^2 \kappa_o^2 - 2\rho_{\hat{z}} \kappa_o \kappa_f + \kappa_f^2 \\ &\Leftrightarrow (\kappa_o - \rho_{\hat{z}} \kappa_f)^2 > (\kappa_f - \rho_{\hat{z}} \kappa_o)^2; \end{aligned} \tag{A.19}$$

so that, $\pi_o^2 > \pi_f^2$ in equation (A.18), proving the third claim.

A.1.3 Proof of Proposition 1.2.4

To prove the second part of the proposition, I first calculate the coefficient ζ_{ri} on $\bar{r}_e$ in the conditional expectation of a_i given s_i, s_e, and $\bar{r}_e$ in equation (1.20). The component of $\bar{r}_e$ orthogonal to s_i and s_e is given by:

$$\hat{r}_e = \bar{r}_e - \mathbb{E}(\bar{r}_e|s_i, s_e) = [a_e - (\delta_o s_e + \delta_f s_i) + \bar{\eta}_e] - [\mu_{ae} - (\delta_o \mu_{se} + \delta_f \mu_{si})], \tag{A.20}$$

where δ_o, δ_f are defined in equation (A.3), and $\bar{\eta}_e$ is the sample mean of $\{\eta_{ue}\}_{u=1}^{t_e}$. Note that the coefficient on $\bar{r}_e$ in the conditional expectation given s_i, s_e, and $\bar{r}_e$ is the same as the coefficient on $\hat{r}_e$ in the conditional expectation given s_i, s_e, and $\hat{r}_e$. Because $\hat{r}_e$ is uncorrelated with s_i and s_e, the coefficient ζ_{ri} is equal to:

$$\zeta_{ri} = \mathbb{C}(a_i, \hat{r}_e)/\mathbb{V}(\hat{r}_e) = \sigma_a^2[\rho_a - \gamma(\delta_o \rho_a + \delta_f)]/(\varsigma^2 + t_e^{-1}\sigma_\eta^2), \tag{A.21}$$

where ς^2 is defined as:

$$\varsigma^2 = \mathbb{V}[a_e - (\delta_o s_e + \delta_f s_i)]. \tag{A.22}$$

Note that the bracketed term in the numerator of equation (A.21) also appears in equation (A.11) and was shown to be positive in the proof of proposition 1.2.2. Thus, ζ_{ri} is positive. Moreover, if $t_1 > t_2$, then $\zeta_{r1} < \zeta_{r2}$.

From equation (1.23), the coefficients ν_{ie}, ν_{ii} in proposition 1.2.4 can be expressed as:

$$\nu_{ii} = (1 - \xi_i)\zeta_{ri}\pi_f + \xi_i \pi_o \;\text{ and }\; \nu_{ie} = (1 - \xi_i)\zeta_{ri}\pi_o + \xi_i \pi_f. \tag{A.23}$$

Thus, the statement $\nu_{12}\nu_{22} < \nu_{21}\nu_{11}$ is equivalent to:

$$[(1-\xi_1)\zeta_{r1}\pi_o + \xi_1\pi_f][(1-\xi_2)\zeta_{r2}\pi_f + \xi_2\pi_o] < [(1-\xi_2)\zeta_{r2}\pi_o + \xi_2\pi_f][(1-\xi_1)\zeta_{r1}\pi_f + \xi_1\pi_o], \tag{A.24}$$

which reduces to:

$$(1 - \xi_1)\xi_2\zeta_{r1}\pi_o^2 + \xi_1(1 - \xi_2)\zeta_{r2}\pi_f^2 < (1 - \xi_2)\xi_1\zeta_{r2}\pi_o^2 + \xi_2(1 - \xi_1)\zeta_{r1}\pi_f^2. \tag{A.25}$$

If $t_1 > t_2$, then $\xi_1 > \xi_2$ and $\zeta_{r1} < \zeta_{r2}$. Thus, equation (A.25) is satisfied if $\pi_o^2 > \pi_f^2$ holds, and proposition 1.2.2 shows that $\pi_o^2 > \pi_f^2$.

A.2 Siblings at Same Age Level

As noted at the end of section 1.2.3, comparing siblings at a given age level instead of in the same time period can provide a more robust test of the individual against the social learning model. When analyzing the wages of siblings at different age levels, one needs to use the coefficient on an individual's own test score to deflate the coefficient on a sibling's test score. This procedure, however, is typically justified only if labor market ability is a unidimensional factor. By instead comparing siblings' wages at the same age level, one avoids having to deflate the regression coefficients in order to test between the two learning models. Nevertheless, a disadvantage of this approach is that younger and older siblings reach the same age at different times, making it necessary to control for other time-varying factors that may affect one's wage.

The proposition below enumerates the predictions of the learning models in sections 1.2.3 and 1.2.4 for the log wages of siblings at a given age level. The first part is an immediate consequence of equations (1.8), (1.9), and (1.12) when the two siblings have the same age t_i. Likewise, the first item in the second part follows directly from equations (1.14), (1.15), and (1.18) whenever $t_1 = t_2$. The results are symmetric if sibling 2 is at least as old as sibling 1 instead of vice versa.

Proposition A.2.1 *Suppose that sibling 1 is at least as old as sibling 2 with $d \geq 0$ being the age difference between them. Let ϑ_{ij} denote the regression coefficient on sibling j's test score in the conditional expectation of sibling i's log wage at age $t > d$ given s_1, s_2 and z_1, z_2.*

1. *If learning is individual, then $\vartheta_{12} = \vartheta_{21}$ and $\vartheta_{11} = \vartheta_{22}$.*

2. *If learning is social, then:*

   (a) $\vartheta_{12} = \vartheta_{21}$ and $\vartheta_{11} = \vartheta_{22}$ for $d = 0$,

   (b) $\vartheta_{12} < \vartheta_{21}$ and $\vartheta_{12}\vartheta_{22} < \vartheta_{21}\vartheta_{11}$ for $d > 0$.

Proof Suppose that sibling 1 is older than sibling 2 by $d > 0$ periods. If employer learning is social, then the market's beliefs about the time-invariant component $g_i(s_i, a_i)$ of sibling i's log labor productivity at age $t > d$ given the variables s_i, s_e and r_i, r_e are normally distributed with mean $\mu_{qi}(s_i, s_e, r_i, r_e)$ and variance σ^2_{qi}, which can be expressed as:

$$\mu_{bi}(s_i, s_e, r_i, r_e) = \beta s_i + \mathbb{E}(a_i|s_i, s_e, r_i, r_e) \;\text{ and }\; \sigma^2_{qi} = \mathbb{V}(a_i|s_i, s_e, r_i, r_e), \tag{A.26}$$

where r_i is a vector of t productivity signals for sibling $i \in \{1, 2\}$, and r_e is a vector of $t - d$ productivity signals for sibling 2 if $i = 1$, $e = 2$ and a vector of $t + d$ productivity signals for sibling 1 if $i = 2$, $e = 1$. Note that, as in section 1.2.4, if i is the index of a given sibling, then e is the index of the other sibling; so that, $e = 1$ if $i = 2$ and vice versa.

Therefore, the conditional expectation of sibling i's labor productivity $l(s_i, a_i, t)$ at age t given s_i, s_e and r_i, r_e is:

$$\mathbb{E}\{\exp[l(s_i, a_i, t)]|s_i, s_e, r_i, r_e\} = \exp[\mu_{bi}(s_i, s_e, r_i, r_e) + \tfrac{1}{2}\sigma^2_{qi} + h(t)], \tag{A.27}$$

which yields the following expression for sibling i's log wage:

$$\log(v_i) = \mu_{bi}(s_i, s_e, r_i, r_e) + \tfrac{1}{2}\sigma^2_{qi} + h(t). \tag{A.28}$$

Because the error terms in each sibling's productivity signals are identically distributed and independent of each other and all of the other variables in the model, the conditional expectation of a_i given s_i, s_e and r_i, r_e can be expressed as:

$$\mathbb{E}(a_i|s_i, s_e, r_i, r_e) = \mathbb{E}(a_i|s_i, s_e, \bar{r}_i, \bar{r}_e) = \varphi_{ci} + \varphi_{hi}s_i + \varphi_{ki}s_e + \varphi_{ai}\bar{r}_i + \varphi_{bi}\bar{r}_e; \tag{A.29}$$

so that, the conditional expectation of sibling i's log wage at age t given s_i, s_e and z_i, z_e has the form:

$$\mathbb{E}[\log(v_i)|s_i, s_e, z_i, z_e] = \varphi_{ai}\mathbb{E}(a_i|s_i, s_e, z_i, z_e) + \varphi_{bi}\mathbb{E}(a_e|s_i, s_e, z_i, z_e) + c_i(s_i, s_e, t). \tag{A.30}$$

It follows from equation (A.30) that the coefficients ϑ_{ii} and ϑ_{ie} in proposition A.2.1 are given by:

$$\vartheta_{ii} = \varphi_{ai}\pi_o + \varphi_{bi}\pi_f \;\text{ and }\; \vartheta_{ie} = \varphi_{ai}\pi_f + \varphi_{bi}\pi_o. \tag{A.31}$$

I now calculate the coefficients φ_{ai} and φ_{bi} on $\bar{r}_i$ and $\bar{r}_e$ in the regression of a_i on s_i, s_e and $\bar{r}_i$, $\bar{r}_e$. The components of $\bar{r}_i$ and $\bar{r}_e$ orthogonal to s_i and s_e are given by:

$$\begin{aligned}\begin{pmatrix}\hat{r}_i\\ \hat{r}_e\end{pmatrix}&=\begin{pmatrix}\bar{r}_i\\ \bar{r}_e\end{pmatrix}-\mathbb{E}\left[\begin{pmatrix}\bar{r}_i\\ \bar{r}_e\end{pmatrix}\mid\begin{pmatrix}s_i\\ s_e\end{pmatrix}\right]\\ &=\begin{pmatrix}a_i-(\delta_o s_i+\delta_f s_e)+\bar{\eta}_i\\ a_e-(\delta_f s_i+\delta_o s_e)+\bar{\eta}_e\end{pmatrix}-\begin{pmatrix}\mu_{ai}-(\delta_o\mu_{si}+\delta_f\mu_{se})\\ \mu_{ae}-(\delta_f\mu_{si}+\delta_o\mu_{se})\end{pmatrix},\end{aligned}\tag{A.32}$$

where δ_o and δ_f are defined in equation (A.3), $\bar{\eta}_i$ is the sample mean of $\{\eta_{ui}\}_{u=1}^{t}$, and $\bar{\eta}_e$ is the sample mean of $\{\eta_{u2}\}_{u=1}^{t-d}$ if $i=1$, $e=2$ and the sample mean of $\{\eta_{u1}\}_{u=1}^{t+d}$ if $i=2$, $e=1$. Note that the coefficients on $\bar{r}_i$ and $\bar{r}_e$ in the regression of a_i on s_i, s_e and $\bar{r}_i$, $\bar{r}_e$ are the same as the coefficients on $\hat{r}_i$ and $\hat{r}_e$ in the regression of a_i on s_i, s_e and $\hat{r}_i$, $\hat{r}_e$. Because $\hat{r}_i$ and $\hat{r}_e$ are uncorrelated with s_i and s_e, the regression coefficients φ_{ai} and φ_{bi} are equal to:

$$\begin{pmatrix}\varphi_{ai} & \varphi_{bi}\end{pmatrix}=\mathbb{C}\left[a_i,\begin{pmatrix}\hat{r}_i\\ \hat{r}_e\end{pmatrix}\right]\mathbb{V}\left[\begin{pmatrix}\hat{r}_i\\ \hat{r}_e\end{pmatrix}\right]^{-1}=\begin{pmatrix}\upsilon_o & \upsilon_f\end{pmatrix}\begin{pmatrix}\upsilon_o+t^{-1}\sigma_\eta^2 & \upsilon_f\\ \upsilon_f & \upsilon_o+t_e^{-1}\sigma_\eta^2\end{pmatrix}^{-1},\tag{A.33}$$

where $t_e=t-d$ if $i=1$, $e=2$ and $t_e=t+d$ if $i=2$, $e=1$, and υ_o and υ_f are given by:

$$\upsilon_o=\sigma_a^2\tau_o=\sigma_a^2[1-\gamma(\delta_o+\delta_f\rho_a)]\ \text{ and }\ \upsilon_f=\sigma_a^2\tau_f=\sigma_a^2[\rho_a-\gamma(\delta_o\rho_a+\delta_f)].\tag{A.34}$$

Note that τ_o and τ_f in the preceding equation also appear in equations (A.7) and (A.11) and were shown to satisfy $\tau_o>\tau_f>0$ in the proof of proposition 1.2.2. Therefore, $\upsilon_o>\upsilon_f>0$. From equation (A.33), the coefficients φ_{ai} and φ_{bi} can be expressed as:

$$\varphi_{ai}=\frac{\upsilon_o(\upsilon_o+t_e^{-1}\sigma_\eta^2)-\upsilon_f^2}{(\upsilon_o+t^{-1}\sigma_\eta^2)(\upsilon_o+t_e^{-1}\sigma_\eta^2)-\upsilon_f^2}\ \text{ and }\ \varphi_{bi}=\frac{\upsilon_f(\upsilon_o+t^{-1}\sigma_\eta^2)-\upsilon_o\upsilon_f}{(\upsilon_o+t^{-1}\sigma_\eta^2)(\upsilon_o+t_e^{-1}\sigma_\eta^2)-\upsilon_f^2}.\tag{A.35}$$

In order to show that $\vartheta_{12}<\vartheta_{21}$, I calculate the derivative of ϑ_{ie} in equation (A.31) with respect to t_e. Using the expressions for φ_{ai} and φ_{bi} in equation (A.35), the derivative can be expressed as:

$$\frac{\partial\vartheta_{ie}}{\partial t_e}=\frac{\upsilon_f\sigma_\eta^4[t(\upsilon_o\pi_o-\upsilon_f\pi_f)+\sigma_\eta^2\pi_o]}{[tt_e(\upsilon_o^2-\upsilon_f^2)+(t+t_e)\upsilon_o\sigma_\eta^2+\sigma_\eta^4]^2}>0,\tag{A.36}$$

where $\upsilon_o > \upsilon_f > 0$ and $\pi_o > \pi_f$. Because $t_e = t - d$ for $i = 1$, $e = 2$ and $t_e = t + d$ for $i = 2$, $e = 1$, one has $\vartheta_{12} < \vartheta_{21}$ as desired.

I now show that $\vartheta_{12}\vartheta_{22} < \vartheta_{21}\vartheta_{11}$. From equation (A.31), one can define:

$$\upsilon = \vartheta_{21}\vartheta_{11} - \vartheta_{12}\vartheta_{22} = (\varphi_{a2}\pi_f + \varphi_{b2}\pi_o)(\varphi_{a1}\pi_o + \varphi_{b1}\pi_f) - (\varphi_{a1}\pi_f + \varphi_{b1}\pi_o)(\varphi_{a2}\pi_o + \varphi_{b2}\pi_f). \quad \text{(A.37)}$$

Using equation (A.35) to substitute for φ_{ai} and φ_{bi}, one obtains after some simplification:

$$\upsilon = \frac{t(\pi_o^2 - \pi_f^2)(t_{e2} - t_{e1})\upsilon_o\upsilon_f\sigma_\eta^4}{[tt_{e1}(\upsilon_o^2 - \upsilon_f^2) + (t + t_{e1})\upsilon_o\sigma_\eta^2 + \sigma_\eta^4][tt_{e2}(\upsilon_o^2 - \upsilon_f^2) + (t + t_{e2})\upsilon_o\sigma_\eta^2 + \sigma_\eta^4]} > 0, \quad \text{(A.38)}$$

where $\pi_o^2 > \pi_f^2$, $t_{e1} = t - d$, $t_{e2} = t + d$, and $\upsilon_o > \upsilon_f > 0$. Because $\upsilon > 0$, one has $\vartheta_{12}\vartheta_{22} < \vartheta_{21}\vartheta_{11}$ as desired. ∎

On the one hand, if the learning process is purely individual, then there will be no differences between siblings in the coefficients obtained from the regression of one's log wage at a given age on one's own and a sibling's schooling and test scores.[1] This outcome also arises if the learning process is social but the siblings have the same age; so that, each sibling has an equally strong impact on the market's beliefs about the other sibling. On the other hand, if there are interactions in the learning process between an older and a younger sibling, then as in the second part of proposition 1.2.4, the ratio of the coefficient on a sibling's test score to the coefficient on one's own test score is typically higher for the younger than for the older sibling in the regression of one's wage at a given age on one's own and a sibling's test scores and schooling. In particular, if sibling 1 is older than sibling 2, then the length of sibling 2's performance history when sibling 1 reaches a given age is less than the length of sibling 1's performance history when sibling 2 reaches that age. Therefore, sibling 2's performance history when sibling 1 is a given age contains less information than sibling 1's performance history when sibling 2 is that age. As a result, sibling 1's wage at a given age places more weight on her own and less on her sibling's performance compared to sibling 2's wage at that age.

The social learning model also makes a definitive prediction regarding the relative values of the coefficients on a sibling's test score. If one sibling is older than the other, then the coefficient on

[1] However, the constant term in the conditional expectation function can still differ between siblings who do not have the same mean values of the regression variables.

a sibling's test score is higher for the younger than for the older sibling in the regression of one's log wage at a given age on one's own and a sibling's test scores and schooling. The logic for this result is that an older sibling's performance is observed over a longer length of time; so that, an older sibling's performance history is a stronger predictor of her own and therefore her sibling's ability. By contrast, it is unclear in general whether the older or the younger sibling has a higher coefficient on her own test score when one's log wage at a given age is regressed on one's own and a sibling's test scores and schooling. That is, if sibling 1 is older than sibling 2, then it is possible under social learning to obtain either $\vartheta_{11} < \vartheta_{22}$ or $\vartheta_{22} > \vartheta_{11}$ from the regression described in proposition A.2.1.

The following two examples illustrate how this ambiguity arises. To understand the case where $\vartheta_{11} < \vartheta_{22}$, suppose that $\rho_a = 1$; so that, the abilities of the two siblings are the same. In this case, the coefficient on one's own test score is equal to the coefficient on a sibling's test score in the regression of one's ability on one's own and a sibling's test scores and schooling. The same is true if one's log wage instead of one's ability is used as the dependent variable in this regression. If sibling 1 is older than sibling 2, then the performance history of sibling 2 when sibling 1 reaches a given age is shorter than the performance history of sibling 1 when sibling 2 reaches that age. This implies that employers have less information about the common ability of the two siblings when sibling 1 is a given age than when sibling 2 is that age. Hence, sibling 1's log wage at a given age is less closely linked to the common ability of the two siblings than sibling 2's log wage at that age. As a result, the coefficient on one's own test score is lower for sibling 1 than for sibling 2; so that, $\vartheta_{11} < \vartheta_{22}$ if $\rho_a = 1$.

To understand the case where $\vartheta_{11} > \vartheta_{22}$, suppose that $\pi_f < 0$; so that, one's ability has a negative partial correlation with a sibling's test score given one's own test score as well as one's own and a sibling's schooling.[2] If sibling 1 is older than sibling 2, then sibling 1 has a higher coefficient on her own performance and a lower coefficient on her sibling's performance in her log wage equation at a given age in comparison to sibling 2 at the same age. That is, sibling 1's log wage has a higher coefficient than sibling 2's log wage on a variable having a positive partial correlation with one's own test score and a lower coefficient than sibling 2's log wage on a variable

[2] See section 1.2.2 for a discussion of the properties of the coefficients π_f and π_o obtained from the regression of one's ability on one's own and a sibling's test scores and schooling.

having a negative partial correlation with one's own test score. Hence, sibling 1 has a higher coefficient than sibling 2 on one's own test score in the regression of one's log wage at a given age level on one's own and a sibling's test scores and schooling; so that, $\vartheta_{11} > \vartheta_{22}$ if $\pi_f < 0$.

A.3 Endogeneity of Test Score

Although the model in section 1.2.1 accounts to some extent for the causal effect of schooling on test scores, the analysis there assumes that one's schooling at test administration is the same as one's schooling when in the labor market. Because individuals can acquire additional schooling after the test is administered, it is important to examine the case where one's schooling at test administration differs from one's schooling when in the labor market. This appendix, therefore, discusses the conditions under which the results in sections 1.2.3 and 1.2.4 remain unchanged if one also controls for schooling at test administration when analyzing the structure of the log wage.

The model analyzed here is the same as that in section 1.2.1, except that equation (1.5) for the test score z_i is replaced with:

$$z_i = \theta_s \tilde{s}_i + \theta_a a_i + \omega_i, \tag{A.39}$$

where $\tilde{s}_i$ denotes one's schooling at the time of taking the test, and all the other variables in the model are defined as before. Letting s_i continue to denote one's schooling when in the labor market, the variables $\tilde{s}_1$ and $\tilde{s}_2$ are assumed to satisfy the following two redundancy conditions:

$$\mathbb{E}(a_i|s_1, s_2, \tilde{s}_1, \tilde{s}_2) = \mathbb{E}(a_i|s_1, s_2) \text{ and } \mathbb{E}(\omega_i|s_1, s_2, \tilde{s}_1, \tilde{s}_2) = \mathbb{E}(\omega_i|s_1, s_2) = 0. \tag{A.40}$$

The restrictions above require one's own and a sibling's schooling at test administration to be uninformative about one's underlying ability a_i and testing error ω_i after controlling for one's own and a sibling's schooling when in the labor market. These assumptions are justifiable if one's schooling when in the labor market represents one's optimal level of schooling given one's ability while one's schooling at test administration is an exogenous lower bound on one's optimal level of schooling. In any case, as explained later in this appendix, these assumptions can to some extent be tested in the available data. Finally, because the test score itself is treated as being unobservable to employers, it is also reasonable to assume that employers do not directly observe an individual's schooling at test administration. As in section 1.2.1, the performance errors η_{1u} and η_{2u} in equation (1.4) are required to be independent of all the other variables in the model, including $\tilde{s}_1$ and $\tilde{s}_2$.

Given the irrelevance conditions in equation (A.40), it is straightforward to show that the predictions of propositions 1.2.3 and 1.2.4 remain unchanged if the test scores z_1 and z_2 are given by equation (A.39) instead of equation (1.5) but the schooling levels at test administration $\tilde{s}_1$ and $\tilde{s}_2$ are added to the existing set of regressors s_1, s_2 and z_1, z_2. In particular, the log wage equations (1.11) and (1.17) are unaffected, because $\tilde{s}_1$ and $\tilde{s}_2$ are unobservable to employers. Furthermore, the restrictions in equation (A.40) imply that the component of z_i orthogonal to s_1, s_2 in the model in section 1.2.2 is the same as the component of z_i orthogonal to s_1, s_2 and $\tilde{s}_1$, $\tilde{s}_2$ in the model in this appendix. Therefore, the regression coefficients π_o and π_f in proposition 1.2.2 are the same as the coefficients on one's own and a sibling's test scores obtained here from the regression of a_i on s_1, s_2 and z_1, z_2 as well as $\tilde{s}_1$, $\tilde{s}_2$. Because the relationship of one's ability to one's own and a sibling's test scores is unchanged conditional on the other control variables, the regression coefficients in propositions 1.2.3 and 1.2.4 are the same as those in the model here when one's own and a sibling's schooling levels at test administration are also included among the set of regressors. The result below summarizes the above discussion.

Proposition A.3.1 *Assume that z_i is given by equation (A.39). Suppose that sibling 1 is at least as old as sibling 2 with $d \geq 0$ being the age difference between them. Let ν_{ij} denote the regression coefficient on sibling j's test score in the linear projection of sibling i's log wage on s_1, s_2 and z_1, z_2 as well as $\tilde{s}_1$, $\tilde{s}_2$.*

1. *If learning is individual, then $\nu_{12}\nu_{22} = \nu_{21}\nu_{11}$.*

2. *If learning is social, then:*

   (a) $\nu_{12}\nu_{22} = \nu_{21}\nu_{11}$ *for* $d = 0$*,*

   (b) $\nu_{12}\nu_{22} < \nu_{21}\nu_{11}$ *for* $d > 0$*.*

There are two possible approaches to testing the restrictions in equation (A.40). First, it follows from equation (A.39) that these restrictions have the following implications for the coefficients obtained from the regression of z_i on s_1, s_2 and $\tilde{s}_1$, $\tilde{s}_2$. The coefficients on one's own schooling both at test administration and when in the labor market should be positive—the former because of the causal effect of schooling on the test score, the latter because of the positive partial correlation between one's schooling and one's ability. Moreover, the coefficient on a sibling's schooling at test administration should be zero due to its irrelevance in predicting any of the components of one's own test score given the other regressors, whereas the coefficient on a sibling's schooling when in the labor market can be either positive or negative as explained in section 1.2.2, depending on whether ability or schooling is more highly correlated between siblings.

Second, information on the log wages of the two siblings can be used to test equation (A.40), especially the restriction on the conditional expectation of each sibling's ability a_i given s_1, s_2 and $\tilde{s}_1$, $\tilde{s}_2$. Equations (1.11) and (1.17) imply that the coefficients on one's own and a sibling's schooling at test administration should both be zero in the regression of one's log wage on s_1, s_2 and $\tilde{s}_1$, $\tilde{s}_2$. This prediction, which is true regardless of whether learning is individual or social, follows because one's log wage can be written as a linear combination of a constant and one's own and a sibling's underlying abilities, performance errors, and schooling levels when in the labor market. However, neither sibling's schooling at test administration is useful in predicting any of these components of the log wage after controlling for s_1 and s_2.

A.4 Sibships of Arbitrary Size

To simplify the exposition, the model developed in section 1.2 emphasizes the case in which each family is composed of only two siblings. However, some households in the data contain three or more interviewed siblings, all of whom may interact with each other in the learning process. It is, therefore, natural to examine the empirical implications of social learning for families with several siblings. This appendix generalizes the learning models in section 1.2 to include an arbitrary number $N \geq 2$ of siblings in each family. The purpose of the analysis in this appendix is to derive the implications of individual and social learning for the coefficients obtained from the regression of one's log wage on one's own and a sibling's test scores after controlling for one's own as well as every sibling's schooling. Overall, the results here demonstrate that the predictions of propositions 1.2.3 and 1.2.4 remain essentially unchanged when considering families with arbitrarily many siblings.

The assumptions made in this appendix about the labor market characteristics of siblings are the same as those in section 1.2.1 with the exception that there are now N instead of two siblings in a given sibship. In particular, the statements in section 1.2.1 regarding a two-sibling family are now treated as holding for every pair of siblings belonging to the same N-sibling family. The siblings are labeled from 1 to N. The variances of the underlying variables a_i, ϵ_i, ω_i, and η_{iu} are assumed to be the same for every sibling, although the means of a_i, ϵ_i, and ω_i are allowed to differ among siblings. Likewise, the sibling correlations in a_i, ϵ_i, and ω_i are assumed to be constant across all pairs of siblings from the same family. The mean of η_{iu} is zero for every individual in each period, and η_{iu} is uncorrelated both over time and across siblings. Each of the N siblings can have a different age t_i; so that, the length of the performance history $\{r_{iu}\}_{u=1}^{t_i}$ can vary among siblings.

The analysis here consists of three parts. First, propositions A.4.1 and A.4.2, which extend propositions 1.2.1 and 1.2.2, characterize the relationship of one's own and a sibling's test scores to one's ability, holding constant one's own and every sibling's schooling. Proposition A.4.3 analyzes the problem of predicting one's ability using the test scores of two of one's siblings in addition to the schooling levels of all the members of one's sibship. Second, proposition A.4.4 like proposition 1.2.3 states the key predictions of the model when employer learning is purely individual. Third, propositions A.4.5 to A.4.8 derive the relevant properties of the log wage under the assumption of social learning. Proposition A.4.9 like proposition 1.2.4 examines how the predictions of the model differ when learning is social instead of individual.

The first step is to derive the component of a person's test score that is orthogonal to both her own and all of her siblings' schooling levels. It is useful to let $x_i = (N-1)^{-1}\sum_{j\neq i} s_j$ denote the mean schooling of person i's siblings. Because s_j and s_k have the same variance as well as identical covariances with s_i and z_i for $i \notin \{j, k\}$, the component of one's test score that is orthogonal to one's own and every sibling's schooling is the same as the component of one's test score that is orthogonal to one's own schooling level and the mean schooling of one's siblings. The result below reports the coefficient obtained from the regression of z_i on s_i and x_i.

Proposition A.4.1 *The regression coefficient of z_i on (s_i, x_i) is given by:*

$$
\begin{aligned}
&\mathbb{C}\left[\begin{pmatrix} z_i \end{pmatrix}, \begin{pmatrix} s_i \\ x_i \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_i \\ x_i \end{pmatrix}\right]^{-1} \\
&= \theta_s \begin{pmatrix} 1 & 0 \end{pmatrix} + \theta_a \frac{\gamma\sigma_a^2}{\sigma_s^2(\sigma_x^2 - \rho_s^2\sigma_s^2)} \begin{pmatrix} \sigma_x^2 - \rho_s\rho_a\sigma_s^2 & (\rho_a - \rho_s)\sigma_s^2 \end{pmatrix}.
\end{aligned}
\tag{A.41}
$$

Proof The conditional expectation of z_i given $(s_i, x_i)'$ is:

$$
\mathbb{E}\left[\begin{pmatrix} z_i \end{pmatrix} | \begin{pmatrix} s_i \\ x_i \end{pmatrix}\right] = \begin{pmatrix} \mu_{zi} \end{pmatrix} + \mathbb{C}\left[\begin{pmatrix} z_i \end{pmatrix}, \begin{pmatrix} s_i \\ x_i \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_i \\ x_i \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} s_i \\ x_i \end{pmatrix} - \begin{pmatrix} \mu_{si} \\ \mu_{xi} \end{pmatrix}\right], \tag{A.42}
$$

where the regression coefficient is given by:

$$
\begin{aligned}
&\mathbb{C}\left[\begin{pmatrix} z_i \end{pmatrix}, \begin{pmatrix} s_i \\ x_i \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_i \\ x_i \end{pmatrix}\right]^{-1} \\
&= \left[\theta_s\sigma_s^2 \begin{pmatrix} 1 & \rho_s \end{pmatrix} + \theta_a\gamma\sigma_a^2 \begin{pmatrix} 1 & \rho_a \end{pmatrix}\right] \begin{pmatrix} \sigma_s^2 & \rho_s\sigma_s^2 \\ \rho_s\sigma_s^2 & \sigma_x^2 \end{pmatrix}^{-1},
\end{aligned}
\tag{A.43}
$$

where $\mathbb{C}(z_i, x_i) = \mathbb{C}(z_i, s_j)$ and $\mathbb{C}(s_i, x_i) = \mathbb{C}(s_i, s_j)$ for all $j \neq i$. Inverting the variance matrix and rearranging terms leads to the formula in equation (A.41). ∎

The component of z_i orthogonal to s_1, s_2, …,s_N can be calculated using the preceding result. The next step is to analyze the relationship of one's ability to one's own and a sibling's test scores after controlling for the schooling levels of all the members of one's sibship. Let p and q index two distinct siblings from the same family. Denoting $s = (s_1, s_2, \dots, s_N)'$ and $z = (z_p, z_q)'$, consider the regression coefficient of sibling p and q's abilities on their own test scores as well as their own and all of their siblings' schooling levels:

$$
\mathbb{C}\left[\begin{pmatrix} a_p \\ a_q \end{pmatrix}, \begin{pmatrix} s \\ z \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s \\ z \end{pmatrix}\right]^{-1} = \begin{pmatrix} \Psi_p & \Pi_o & \Pi_f \\ \Psi_q & \Pi_f & \Pi_o \end{pmatrix}. \tag{A.44}
$$

The result below characterizes the regression parameters Π_o and Π_f, which represent the relationship of one's ability to one's own test score and a sibling's test score, holding constant the schooling of all the members of one's sibship.

Proposition A.4.2 *The regression coefficients* Π_o *and* Π_f *satisfy* $\Pi_o > \Pi_f$, $\Pi_o > 0$, *and* $\Pi_o^2 > \Pi_f^2$.

Proof Expressing the regression coefficient in equation (A.41) as:

$$\mathbb{C}\left[\begin{pmatrix} z_i \end{pmatrix}, \begin{pmatrix} s_i \\ x_i \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s_i \\ x_i \end{pmatrix}\right]^{-1} = \begin{pmatrix} \theta_s + \theta_a \Delta_o & \theta_a \Delta_f \end{pmatrix}, \tag{A.45}$$

the component of z_i orthogonal to s, which is the same as the component of z_i orthogonal to s_i and x_i, is given by:

$$\hat{z}_i = z_i - \mathbb{E}(z_i|s_i, x_i) = \{\theta_a[a_i - (\Delta_o s_i + \Delta_f x_i)] + \omega_i\} - \{\theta_a[\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})]\}, \tag{A.46}$$

where equations (1.5) and (A.42) are used to substitute for z_i and $\mathbb{E}(z_i|s_i, x_i)$, respectively. Note that the coefficient on $(z_p, z_q)'$ in a regression on $(s_1, s_2, \ldots, s_N, z_p, z_q)'$ is the same as the coefficient on $(\hat{z}_p, \hat{z}_q)'$ in a regression on $(s_1, s_2, \ldots, s_N, \hat{z}_p, \hat{z}_q)'$. Therefore, consider the regression of $(a_p, a_q)'$ on $(s_1, s_2, \ldots, s_N, \hat{z}_p, \hat{z}_q)'$. Because $(\hat{z}_p, \hat{z}_q)'$ is uncorrelated with $(s_1, s_2, \ldots, s_N)'$, the coefficient on $(\hat{z}_p, \hat{z}_q)'$ can be expressed as:

$$\begin{pmatrix} \Pi_o & \Pi_f \\ \Pi_f & \Pi_o \end{pmatrix} = \mathbb{C}\left[\begin{pmatrix} a_p \\ a_q \end{pmatrix}, \begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right]^{-1}, \tag{A.47}$$

where the inverse variance matrix has the form:

$$\mathbb{V}\left[\begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right]^{-1} = [\sigma_{\hat{z}}^4(1-\rho_{\hat{z}}^2)]^{-1} \begin{pmatrix} \sigma_{\hat{z}}^2 & -\rho_{\hat{z}}\sigma_{\hat{z}}^2 \\ -\rho_{\hat{z}}\sigma_{\hat{z}}^2 & \sigma_{\hat{z}}^2 \end{pmatrix}. \tag{A.48}$$

I first calculate the covariances $\mathbb{C}(a_p, \hat{z}_p) = \mathbb{C}(a_q, \hat{z}_q)$ and $\mathbb{C}(a_p, \hat{z}_q) = \mathbb{C}(a_q, \hat{z}_p)$. Using equations (A.41) and (A.46), the former covariance is equal to:

$$\begin{aligned}\mathbb{C}(a_p, \hat{z}_p) = \mathbb{C}(a_q, \hat{z}_q) &= \theta_a \sigma_a^2 [1 - \gamma(\Delta_o + \Delta_f \rho_a)] \\ &= \theta_a \sigma_a^2 \left(1 - \frac{\gamma_a^2 \sigma_a^2 [\sigma_x^2 - \sigma_s^2 (2\rho_a \rho_s - \rho_a^2)]}{\sigma_s^2 (\sigma_x^2 - \rho_s^2 \sigma_s^2)} \right),\end{aligned} \tag{A.49}$$

and the latter covariance is equal to:

$$\begin{aligned}\mathbb{C}(a_p, \hat{z}_q) = \mathbb{C}(a_q, \hat{z}_p) &= \theta_a \sigma_a^2 [\rho_a - \gamma(\Delta_o \rho_a + \Delta_f \varrho_a)] \\ &= \theta_a \sigma_a^2 \left(\rho_a - \frac{\gamma^2 \sigma_a^2 \{(\sigma_x^2 - \rho_a \rho_s \sigma_s^2)\rho_a + \sigma_s^2 (\rho_a - \rho_s)\varrho_a\}}{\sigma_s^2 (\sigma_x^2 - \rho_s^2 \sigma_s^2)} \right),\end{aligned} \tag{A.50}$$

where the parameter ϱ_a is defined as:

$$\varrho_a = [1 + (N-2)\rho_a]/(N-1). \tag{A.51}$$

I next show that $\mathbb{C}(a_p, \hat{z}_q) > 0$. From equation (A.50), the statement $\mathbb{C}(a_p, \hat{z}_q) > 0$ is equivalent to:

$$\begin{aligned}K =& \sigma_s^2 \rho_a [1 + (N-2)\rho_s - (N-1)\rho_s^2] \\ &- \gamma^2 \sigma_a^2 \{\rho_a [1 + (N-2)\rho_s - (N-1)\rho_a \rho_s] + (\rho_a - \rho_s)[1 + (N-2)\rho_a]\} > 0,\end{aligned} \tag{A.52}$$

where the variance σ_s^2 and the correlation ρ_s are given by:

$$\sigma_s^2 = \gamma^2 \sigma_a^2 + \sigma_\epsilon^2 \;\text{ and }\; \rho_s = (\gamma^2 \rho_a \sigma_a^2 + \rho_\epsilon \sigma_\epsilon^2)/(\gamma^2 \sigma_a^2 + \sigma_\epsilon^2). \tag{A.53}$$

Note that $K = 0$ if $\sigma_\epsilon^2 = 0$. Using equation (A.53) to substitute for σ_s^2 and ρ_s in equation (A.52), the derivative of K with respect to σ_ϵ^2 is:

$$\begin{aligned}&\frac{\partial K}{\partial \sigma_\epsilon^2} = \\ &\frac{\sigma_\epsilon^4 \rho_a (1-\rho_\epsilon)[1+(N-1)\rho_\epsilon] + 2\gamma^2 \sigma_a^2 \sigma_\epsilon^2 \rho_a (1-\rho_\epsilon)[1+(N-1)\rho_\epsilon] + \gamma^4 \sigma_a^4 \rho_\epsilon (1-\rho_a)[1+(N-1)\rho_a]}{(\gamma^2 \sigma_a^2 + \sigma_\epsilon^2)^2} > 0.\end{aligned} \tag{A.54}$$

It follows that $K > 0$ if $\sigma_\epsilon^2 > 0$ and so $\mathbb{C}(a_p, \hat{z}_q) > 0$.

I now show that $\mathbb{C}(a_p, \hat{z}_p) > \mathbb{C}(a_p, \hat{z}_q)$. From equations (A.49) and (A.50), the statement $\mathbb{C}(a_p, \hat{z}_p) > \mathbb{C}(a_p, \hat{z}_q)$ is equivalent to:

$$\begin{aligned} W = {} & \sigma_s^2(1-\rho_a)[1+(N-2)\rho_s-(N-1)\rho_s^2] \\ & -\gamma^2\sigma_a^2[1+(N-2)\rho_s-2\rho_s\rho_a(N-1)+(N-1)\rho_a^2] \\ & +\gamma^2\sigma_a^2\rho_a[1+(N-2)\rho_s-(N-1)\rho_s\rho_a]+\gamma^2\sigma_a^2(\rho_a-\rho_s)[1+(N-2)\rho_a] > 0, \end{aligned} \tag{A.55}$$

where σ_s^2 and ρ_s are given by equation (A.53). Note that $W = 0$ if $\sigma_\epsilon^2 = 0$. Using equation (A.53) to substitute for σ_s^2 and ρ_s in equation (A.55), the derivative of W with respect to σ_ϵ^2 is:

$$\begin{aligned} & \frac{\partial W}{\partial \sigma_\epsilon^2} = \\ & \frac{(1-\rho_a)(1-\rho_\epsilon)\{\sigma_\epsilon^4[1+(N-1)\rho_\epsilon]+2\gamma^2\sigma_a^2\sigma_\epsilon^2[1+(N-1)\rho_\epsilon]+\gamma^4\sigma_a^4[1+(N-1)\rho_a]\}}{(\gamma^2\sigma_a^2+\sigma_\epsilon^2)^2} > 0. \end{aligned} \tag{A.56}$$

It follows that $W > 0$ if $\sigma_\epsilon^2 > 0$ and so $\mathbb{C}(a_p, \hat{z}_p) > \mathbb{C}(a_p, \hat{z}_q)$.

Finally, I use the results shown above to prove the three claims in proposition A.4.2. Given the form of the inverse variance matrix in equation (A.48), it follows from $\mathbb{C}(a_p, \hat{z}_p) > \mathbb{C}(a_p, \hat{z}_q)$ that $\Pi_o > \Pi_f$, proving the first claim. From equations (A.47) and (A.48), the regression parameters Π_o and Π_f take the form:

$$\Pi_o = \Gamma_o - \rho_{\hat{z}}\Gamma_f \quad \text{and} \quad \Pi_f = \Gamma_f - \rho_{\hat{z}}\Gamma_o, \tag{A.57}$$

where $\Gamma_o = \tau\mathbb{C}(a_p, \hat{z}_p) > 0$, $\Gamma_f = \tau\mathbb{C}(a_p, \hat{z}_q) > 0$, and $\tau > 0$. Because it has been shown above that $\mathbb{C}(a_p, \hat{z}_p) > \mathbb{C}(a_p, \hat{z}_q) > 0$, one has $\Gamma_o > \Gamma_f > 0$. These inequalities imply that $\Pi_o > 0$ in equation (A.57), proving the second claim. In addition, because $\Gamma_o^2 > \Gamma_f^2$ from the preceding inequalities, one has:

$$\begin{aligned} & \Gamma_o^2+\rho_{\hat{z}}^2\Gamma_f^2 > \rho_{\hat{z}}^2\Gamma_o^2+\Gamma_f^2 \Leftrightarrow \Gamma_o^2-2\rho_{\hat{z}}\Gamma_o\Gamma_f+\rho_{\hat{z}}^2\Gamma_f^2 > \rho_{\hat{z}}^2\Gamma_o^2-2\rho_{\hat{z}}\Gamma_o\Gamma_f+\Gamma_f^2 \\ & \Leftrightarrow (\Gamma_o-\rho_{\hat{z}}\Gamma_f)^2 > (\Gamma_f-\rho_{\hat{z}}\Gamma_o)^2; \end{aligned} \tag{A.58}$$

so that, $\Pi_o^2 > \Pi_f^2$ in equation (A.57), proving the third claim. ∎

Because families can include more than two siblings in the model studied in this appendix, it is also necessary to examine the relationship of sibling p and q's test scores z_p and z_q to the ability a_r

of a third sibling r, holding constant the schooling levels $\{s_k\}_{k=1}^N$ of all the members of the sibship. The regression coefficient for this prediction problem is given by:

$$\mathbb{C}\left[\begin{pmatrix} a_r \end{pmatrix}, \begin{pmatrix} s \\ z \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} s \\ z \end{pmatrix}\right]^{-1} = \begin{pmatrix} \Psi_x & \Pi_x & \Pi_x \end{pmatrix}. \tag{A.59}$$

The result below shows that an individual's ability has a positive partial correlation with a sibling's test score given another sibling's test score as well as her own and all of her siblings' schooling levels.

Proposition A.4.3 *Let $N \geq 3$. Consider three distinct siblings p, q, and r from the same family. Let Π_x denote the regression coefficients on both sibling p and q's test scores z_p and z_q in the conditional expectation of sibling r's ability a_r given z_p, z_q, and $\{s_k\}_{k=1}^N$. Then $\Pi_x > 0$.*

Proof Let $\hat{z}_i$ denote the component of z_i orthogonal to $\{s_k\}_{k=1}^N$. Recall that $\hat{z}_i$ is given by equation (A.46). Note that the coefficient on (z_p, z_q) in a regression on $(s_1, s_2, \ldots, s_N, z_p, z_q)$ is the same as the coefficient on $(\hat{z}_p, \hat{z}_q)$ in a regression on $(\hat{z}_p, \hat{z}_q)$. Therefore, the regression coefficient Π_x can be expressed as:

$$\begin{pmatrix} \Pi_x & \Pi_x \end{pmatrix} = \mathbb{C}\left[\begin{pmatrix} a_r \end{pmatrix}, \begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right]^{-1}, \tag{A.60}$$

where the inverse variance matrix has the form:

$$\mathbb{V}\left[\begin{pmatrix} \hat{z}_p \\ \hat{z}_q \end{pmatrix}\right]^{-1} = [\sigma_{\hat{z}}^4(1-\rho_{\hat{z}}^2)]^{-1} \begin{pmatrix} \sigma_{\hat{z}}^2 & -\rho_{\hat{z}}\sigma_{\hat{z}}^2 \\ -\rho_{\hat{z}}\sigma_{\hat{z}}^2 & \sigma_{\hat{z}}^2 \end{pmatrix}, \tag{A.61}$$

which is analogous to equation (A.48). As in equation (A.50), the covariance terms in equation (A.60) are given by:

$$\Xi_{\hat{z}} = \mathbb{C}(a_r, \hat{z}_p) = \mathbb{C}(a_r, \hat{z}_q) = \theta_a\sigma_a^2\big(\rho_a - \gamma\{\Delta_o\rho_a + \Delta_f[1+(N-2)\rho_a]/(N-1)\}\big), \tag{A.62}$$

which was shown to be positive in the proof of proposition A.4.2. From equations (A.60) and (A.61), the coefficient Π_x can be expressed as:

$$\Pi_x = \Xi_{\hat{z}}[\sigma_{\hat{z}}^2(1+\rho_{\hat{z}})]^{-1}, \tag{A.63}$$

which is positive because $\Xi_{\hat{z}} > 0$. ∎

Before analyzing social interactions among siblings, I consider the benchmark case in which learning is individual. Because only one's own characteristics are used to set wages, an individual's log wage is given as in section 1.2.3 by equation (1.11). Therefore, the conditional expectation of the log wage for person $i \in \{p, q\}$ given z_p, z_q and $s_1, s_2, \ldots, s_N$ can be expressed as:

$$\mathbb{E}[\log(w_i)|s_1, s_2, \ldots, s_N, z_p, z_q] = \chi_i \mathbb{E}(a_i|s_1, s_2, \ldots, s_N, z_p, z_q) + f_i(s_i, t_i), \tag{A.64}$$

where the function $f_i(s_i, t_i)$ is defined in equation (1.13). The following analogue of proposition 1.2.3 is an immediate consequence of equation (A.64). As stated below, if learning is individual, then the ratio of the coefficient on a sibling's test score to the coefficient on one's own test score will be the same for siblings p and q, irrespective of the age of any of the siblings in their family.

Proposition A.4.4 *Suppose that learning is individual. Consider two distinct siblings p and q from the same family. Let α_{ij} denote the regression coefficient on sibling j's test score in the conditional expectation of sibling i's log wage given z_p, z_q and s_1, $s_2, \ldots,$ s_N. Then $\alpha_{pq}\alpha_{qq} = \alpha_{qp}\alpha_{pp} = (\chi_p\Pi_f)(\chi_q\Pi_o)$.*

The remainder of this appendix examines the case in which there are social interactions in the learning process among the N siblings in a family. In particular, each individual's wage is assumed to be equal to the conditional expectation of her own labor productivity given her own and all of her siblings' schooling and performance. Conditional on the schooling $\{s_j\}_{j=1}^N$ and the performance $\{r_j\}_{j=1}^N$ of all the members of a sibship, the market's beliefs about the time-invariant component $g_i(s_i, a_i)$ of sibling i's log labor productivity are normally distributed with mean $\mu_{bi}\big(\{s_j\}_{j=1}^N, \{r_j\}_{j=1}^N\big)$ and variance σ_{bi}^2 given by:

$$\mu_{bi}\big(\{s_j\}_{j=1}^N, \{r_j\}_{j=1}^N\big) = \beta s_i + \mathbb{E}\big(a_i|\{s_j\}_{j=1}^N, \{r_j\}_{j=1}^N\big) \text{ and } \sigma_{bi}^2 = \mathbb{V}\big(a_i|\{s_j\}_{j=1}^N, \{r_j\}_{j=1}^N\big). \tag{A.65}$$

Noting that the error terms in each individual's productivity signals are identically distributed and independent of each other and all of the other variables in the model, the conditional expectation of a_i given $\{s_j\}_{j=1}^N$ and $\{r_j\}_{j=1}^N$ can be expressed as:

$$\mathbb{E}\big(a_i|\{s_j\}_{j=1}^N,\{r_j\}_{j=1}^N\big)=\mathbb{E}\big(a_i|\{s_j\}_{j=1}^N,\{\bar{r}_j\}_{j=1}^N\big)=\gamma_i+\sum_{j=1}^N\delta_{ij}s_j+\sum_{j=1}^N\lambda_{ij}\bar{r}_j, \tag{A.66}$$

where $\bar{r}_j$ is the sample mean of $\{r_{ju}\}_{u=1}^{t_j}$. Because of the normality of the market's beliefs, the conditional expectation of sibling i's labor productivity $l(s_i,a_i,t_i)$ given $\{s_j\}_{j=1}^N$ and $\{r_j\}_{j=1}^N$ is:

$$\mathbb{E}\big(\exp[l(s_i,a_i,t_i)]|\{s_j\}_{j=1}^N,\{r_j\}_{j=1}^N\big)=\exp\big[\mu_{bi}\big(\{s_j\}_{j=1}^N,\{r_j\}_{j=1}^N\big)+\tfrac{1}{2}\sigma_{bi}^2+h(t_i)\big], \tag{A.67}$$

which yields the following expression for sibling i's log wage:

$$\log(v_i)=\mu_{bi}\big(\{s_j\}_{j=1}^N,\{r_j\}_{j=1}^N\big)+\tfrac{1}{2}\sigma_{bi}^2+h(t_i). \tag{A.68}$$

Using equations (A.66) and (A.67) to substitute for $\mu_{bi}\big(\{s_j\}_{j=1}^N,\{r_j\}_{j=1}^N\big)$ in equation (A.68), one obtains:

$$\log(v_i)=\sum_{j=1}^N\lambda_{ij}\bar{r}_j+y_i\big(\{s_j\}_{j=1}^N,t_i\big), \tag{A.69}$$

where the function $y_i\big(\{s_j\}_{j=1}^N,t_i\big)$ is defined as:

$$y_i\big(\{s_j\}_{j=1}^N,t_i\big)=\gamma_i+\beta s_i+\sum_{j=1}^N\delta_{ij}s_j+\tfrac{1}{2}\sigma_{bi}^2+h(t_i). \tag{A.70}$$

In order to understand the behavior of the log wage under social learning, it is necessary to derive the properties of the coefficient λ_{ij} on each sibling j's mean performance $\bar{r}_j$ in the regression of one's ability a_i on $\{s_j\}_{j=1}^N$ and $\{\bar{r}_j\}_{j=1}^N$. Let $\hat{r}_j$ denote the component of $\bar{r}_j$ that is orthogonal to $\{s_j\}_{j=1}^N$. Let $\hat{r}$ represent the vector $(\hat{r}_1,\hat{r}_2,\dots,\hat{r}_N)'$. Then the regression parameter $\lambda_i=(\lambda_{i1},\lambda_{i2},\dots,\lambda_{iN})$ is given by:

$$\lambda_i=\mathbb{C}(a_i,\hat{r})\mathbb{V}(\hat{r})^{-1}. \tag{A.71}$$

The next result establishes a few simple facts about the covariance vector $\mathbb{C}(a_i,\hat{r})$ and the variance matrix $\mathbb{V}(\hat{r})$ that are useful in deriving the properties of the regression coefficient λ_i. First, it shows that all the elements of the covariance matrix $\mathbb{C}(a_i, \hat{r})$ are equal to $\Xi_f > 0$, except for the i^{th} element, which is equal to $\Xi_o > \Xi_f$. Second, it shows that all the off-diagonal elements of the variance matrix $\mathbb{V}(\hat{r})$ are equal to $\Sigma_0 > 0$ but that the i^{th} diagonal element of $\mathbb{V}(\hat{r})$ is equal to $\Sigma_i > \Sigma_0$. Observe that all the entries of $\mathbb{C}(a_i, \hat{r})$, except for the i^{th} one, are equal to each other. Likewise, all the off-diagonal entries of $\mathbb{V}(\hat{r})$ are the same, whereas the diagonal entries can differ, depending on the ages of the siblings.

Proposition A.4.5 *For any i, let $\Xi_o = \mathbb{C}(a_i, \hat{r}_i)$ and $\Sigma_i = \mathbb{V}(\hat{r}_i)$. For any i, j such that $i \neq j$, let $\Xi_f = \mathbb{C}(a_i, \hat{r}_j)$ and $\Sigma_0 = \mathbb{C}(\hat{r}_i, \hat{r}_j)$. Then the vector $\mathbb{C}(a_i, \hat{r})$ satisfies $\Xi_o > \Xi_f > 0$, and the invertible matrix $\mathbb{V}(\hat{r})$ satisfies $\Sigma_i > \Sigma_0 > 0$.*

Proof I first calculate the covariance $\mathbb{C}(a_i, \hat{r}_k)$ both for $k = i$ and for $k \neq i$. The component of $\bar{r}_k$ orthogonal to $\{s_k\}_{k=1}^N$, which is identical to the component of $\bar{r}_k$ orthogonal to s_k and x_k, is given by:

$$\hat{r}_k = \bar{r}_k - \mathbb{E}(\bar{r}_k | s_k, x_k) = [a_k - (\Delta_o s_k + \Delta_f x_k) + \bar{\eta}_k] - [\mu_{ak} - (\Delta_o \mu_{sk} + \Delta_f \mu_{xk})], \tag{A.72}$$

where Δ_o and Δ_f are defined in equation (A.45), and $\bar{\eta}_k$ is the sample mean of $\{\eta_{uk}\}_{u=1}^{t_k}$. Therefore, the covariance $\mathbb{C}(a_i, \hat{r}_k)$ is as follows:

$$\mathbb{C}(a_i, \hat{r}_k) = \begin{cases} \sigma_a^2[1 - \gamma(\Delta_o + \Delta_f \rho_a)] & \text{if } k = i \\ \sigma_a^2[\rho_a - \gamma(\Delta_o \rho_a + \Delta_f \varrho_a)] & \text{if } k \neq i \end{cases}, \tag{A.73}$$

where ϱ_a is defined in equation (A.51). The bracketed terms in equation (A.73), which also appear in equations (A.49) and (A.50), were shown in the proof of proposition A.4.2 to satisfy the following:

$$1 - \gamma(\Delta_o + \Delta_f \rho_a) > \rho_a - \gamma(\Delta_o \rho_a + \Delta_f \varrho_a) > 0. \tag{A.74}$$

Therefore, one has $\mathbb{C}(a_i, \hat{r}_i) > \mathbb{C}(a_i, \hat{r}_j) > 0$ for all $j \neq i$, which is equivalent to $\Xi_o > \Xi_f > 0$ as desired.

I next calculate the covariance $\mathbb{C}(\hat{r}_i, \hat{r}_j)$ for all i, j such that $i \neq j$. Observe that $\mathbb{C}(\hat{r}_i, \hat{r}_j)$

reduces to:

$$\mathbb{C}(\hat{r}_i, \hat{r}_j) = \mathbb{C}([a_i - (\Delta_o s_i + \Delta_f x_i) + \bar{\eta}_i] - [\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})], \hat{r}_j) = \mathbb{C}(a_i, \hat{r}_j), \quad \text{(A.75)}$$

where the first step follows from substituting for $\hat{r}_i$ using equation (A.72), and the second step uses the fact that $\hat{r}_j$ has zero covariance with s_i and x_i. I now compute the variance $\mathbb{V}(\hat{r}_i)$ for all i. Note that $\mathbb{V}(\hat{r}_i)$ can be expressed as:

$$\begin{aligned} \mathbb{V}(\hat{r}_i) &= \mathbb{C}([a_i - (\Delta_o s_i + \Delta_f x_i) + \bar{\eta}_i] - [\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})], \hat{r}_i) \\ &= \mathbb{C}(a_i, \hat{r}_i) + \mathbb{C}(\bar{\eta}_i, \hat{r}_i) = \mathbb{C}(a_i, \hat{r}_i) + t_i^{-1}\sigma_\eta^2, \end{aligned} \quad \text{(A.76)}$$

where the first step substitutes for $\hat{r}_i$ from equation (A.72), the second step holds because $\hat{r}_i$ is orthogonal to s_i and x_i, and the third step follows because $\bar{\eta}_i$ has a nonzero covariance only with itself. Because it has been shown above that $\mathbb{C}(a_i, \hat{r}_i) > \mathbb{C}(a_i, \hat{r}_j) > 0$ for all $j \neq i$, the two preceding results $\mathbb{C}(\hat{r}_i, \hat{r}_j) = \mathbb{C}(a_i, \hat{r}_j)$ and $\mathbb{V}(\hat{r}_i) = \mathbb{C}(a_i, \hat{r}_i) + t_i^{-1}\sigma_\eta^2$ imply that $\mathbb{V}(\hat{r}_i) > \mathbb{C}(\hat{r}_i, \hat{r}_j) > 0$ for all $j \neq i$, which is equivalent to $\Sigma_i > \Sigma_0 > 0$ as desired.

I finally show that the matrix $\mathbb{V}(\hat{r})$ is invertible. Note that $\mathbb{V}(\hat{r})$ can be decomposed as:

$$\mathbb{V}(\hat{r}) = A + uu', \quad \text{(A.77)}$$

where A is the $N \times N$ diagonal matrix whose k^{th} diagonal entry is $\Sigma_k - \Sigma_0$, and u is an $N \times 1$ vector each of whose entries is $\sqrt{\Sigma_0}$. Then the determinant of $\mathbb{V}(\hat{r})$ can be calculated as follows:

$$\begin{aligned} \det[\mathbb{V}(\hat{r})] = &\det(A + uu') = (1 + u'A^{-1}u)\det(A) = \\ &[1 + \sum_{k=1}^{N} \Sigma_0(\Sigma_k - \Sigma_0)^{-1}] \prod_{k=1}^{N} (\Sigma_k - \Sigma_0) > 0, \end{aligned} \quad \text{(A.78)}$$

where the second equality follows from the matrix determinant lemma, and the inequality follows because $\Sigma_k > \Sigma_0 > 0$ for all k as shown above. Since $\det[\mathbb{V}(\hat{r})] > 0$, $\mathbb{V}(\hat{r})$ is invertible. ∎

The behavior of the regression coefficient λ_i depends on the structure of the inverse variance matrix $\mathbb{V}(\hat{r})^{-1}$ in equation (A.71). Therefore, it is necessary to obtain a precise characterization of the inverse of the $N \times N$ matrix $\mathbb{V}(\hat{r})$ before proceeding with the analysis of the coefficient λ_i. The result below describes the relevant features of the inverses of matrices having a structure similar to $\mathbb{V}(\hat{r})$. In particular, I consider an arbitrary $\tilde{N} \times \tilde{N}$ matrix $\tilde{V}$ all of whose off-diagonal elements are equal to the same positive constant and each of whose diagonal elements is greater than any off-diagonal element. The result below shows that the matrix inverse $\tilde{V}^{-1}$ has positive diagonal entries and negative off-diagonal entries. Moreover, the sum of the entries in each column of $\tilde{V}^{-1}$ is shown to be positive.

Proposition A.4.6 *Let $\tilde{V}$ be an arbitrary $\tilde{N} \times \tilde{N}$ matrix whose off-diagonal entries are all equal to $\tilde{\Sigma}_0 > 0$ and whose i^{th} diagonal entry is equal to $\tilde{\Sigma}_i > \tilde{\Sigma}_0$. Let $\tilde{\Lambda}_{ij}$ denote the $(i, j)^{th}$ element of the matrix inverse $\tilde{V}^{-1}$. Then:*

1. *$\tilde{\Lambda}_{ii} > 0$ for all i,*

2. *$\tilde{\Lambda}_{ij} < 0$ for $i \neq j$,*

3. *$\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kj} > 0$ for all j.*

Proof From the definition of a matrix inverse, the parameters $\tilde{\Lambda}_{ij}$ must satisfy the following equation for all i, j:

$$\tilde{\Sigma}_i \tilde{\Lambda}_{ij} + \tilde{\Sigma}_0 \sum_{k \neq i} \tilde{\Lambda}_{kj} = \delta_{ij}, \tag{A.79}$$

where δ_{ij} is the Kronecker delta equal to 1 if $i = j$ and 0 if $i \neq j$. Recall from the statement of the proposition that $\tilde{\Sigma}_i > \tilde{\Sigma}_0 > 0$ for all i.

I first show that the sum $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kj}$ of the elements in each column j of the matrix $\tilde{V}^{-1}$ is non-negative. Suppose to the contrary that $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kl}$ is negative for some l. Then there must be some m such that $\tilde{\Lambda}_{ml}$ is negative. For this pair of l and m, it follows that:

$$0 > \tilde{\Sigma}_0 \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m} \tilde{\Lambda}_{kl} > \tilde{\Sigma}_m \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m} \tilde{\Lambda}_{kl}, \tag{A.80}$$

where the second step holds because $\tilde{\Lambda}_{ml} < 0$ and $\tilde{\Sigma}_m > \tilde{\Sigma}_0 > 0$. However, the preceding equation contradicts equation (A.79), which requires that $\tilde{\Sigma}_m \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m}^{\tilde{N}} \tilde{\Lambda}_{kl} \geq 0$. Thus, $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kl}$ must be non-negative.

I next show that the off-diagonal elements of the matrix $\tilde{V}^{-1}$ are non-positive; that is, $\tilde{\Lambda}_{ij} \leq 0$ for $i \neq j$. Suppose to the contrary that $\tilde{\Lambda}_{ab} > 0$ for some a and b such that $a \neq b$. From equation (A.79), one has:

$$\tilde{\Sigma}_a \tilde{\Lambda}_{ab} + \tilde{\Sigma}_0 \sum_{k \neq a} \tilde{\Lambda}_{kb} = 0. \tag{A.81}$$

Because $\tilde{\Sigma}_a > \tilde{\Sigma}_0 > 0$, the preceding equation implies:

$$\tilde{\Sigma}_0 \tilde{\Lambda}_{ab} + \tilde{\Sigma}_0 \sum_{k \neq a} \tilde{\Lambda}_{kb} < 0 \quad \Leftrightarrow \quad \sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kb} < 0, \tag{A.82}$$

which contradicts the result shown above that the sum $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kj}$ of each column j of the matrix $\tilde{V}^{-1}$ must be non-negative. Thus, $\tilde{\Lambda}_{ij} \leq 0$ for $i \neq j$.

I now show that the diagonal elements of the matrix $\tilde{V}^{-1}$ are positive; that is, $\tilde{\Lambda}_{ii} > 0$ for all i. From equation (A.79), one has:

$$\tilde{\Sigma}_i \tilde{\Lambda}_{ii} + \tilde{\Sigma}_0 \sum_{k \neq i} \tilde{\Lambda}_{ki} = 1. \tag{A.83}$$

Because it has been shown above that $\tilde{\Lambda}_{ij} \leq 0$ for $i \neq j$, it must be that $\sum_{k \neq i} \tilde{\Lambda}_{ki} \leq 0$. Therefore, noting that $\tilde{\Sigma}_i > 0$ and $\tilde{\Sigma}_0 > 0$, one must have $\tilde{\Lambda}_{ii} > 0$, in order for the preceding equation to hold. This proves the first claim.

I also show that the sum $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kj}$ of the elements in each column j of the matrix $\tilde{V}^{-1}$ is positive. Suppose to the contrary that $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kl}$ is non-positive for some l. Because it has been shown above that $\tilde{\Lambda}_{ii} > 0$ for all i, there must be some m with $m \neq l$ such that $\tilde{\Lambda}_{ml}$ is negative. For this pair of l and m, it follows that:

$$0 \geq \tilde{\Sigma}_0 \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m} \tilde{\Lambda}_{kl} > \tilde{\Sigma}_m \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m} \tilde{\Lambda}_{kl}, \tag{A.84}$$

where the second step follows because $\tilde{\Lambda}_{ml} < 0$ and $\tilde{\Sigma}_m > \tilde{\Sigma}_0 > 0$. However, equation (A.79) indicates that $\tilde{\Sigma}_m \tilde{\Lambda}_{ml} + \tilde{\Sigma}_0 \sum_{k \neq m} \tilde{\Lambda}_{kl} = 0$, which contradicts the preceding equation. Thus, $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kl}$ must be positive, which proves the third claim.

I finally show that the off-diagonal elements of the matrix $\tilde{V}^{-1}$ are negative; that is, $\tilde{\Lambda}_{ij} < 0$ for $i \neq j$. Suppose to the contrary that $\tilde{\Lambda}_{ab} \geq 0$ for some a and b such that $a \neq b$. From equation

(A.79), one has:

$$\tilde{\Sigma}_a\tilde{\Lambda}_{ab} + \tilde{\Sigma}_0 \sum_{k\neq a} \tilde{\Lambda}_{kb} = 0. \tag{A.85}$$

Because $\tilde{\Sigma}_a > \tilde{\Sigma}_0 > 0$, the preceding equation implies:

$$\tilde{\Sigma}_0\tilde{\Lambda}_{ab} + \tilde{\Sigma}_0 \sum_{k\neq a} \tilde{\Lambda}_{kb} \leq 0 \quad \Leftrightarrow \quad \sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kb} \leq 0, \tag{A.86}$$

which contradicts the result shown above that the sum $\sum_{k=1}^{\tilde{N}} \tilde{\Lambda}_{kj}$ of each column j of the matrix $\tilde{V}^{-1}$ must be positive. Thus, $\tilde{\Lambda}_{ij} < 0$ for $i \neq j$, proving the second claim. ∎

It is now possible to examine the inference problem facing employers with imperfect information about worker ability. The next result shows that each person j's average performance has a positive coefficient λ_{ij} in person i's log wage, which is given by equation (A.69). That is, an individual's wage assigns a positive price to both her own and every sibling's performance.

Proposition A.4.7 *Suppose that learning is social. Let λ_{ij} denote the coefficient on sibling j's average performance $\bar{r}_j$ in the regression of sibling i's ability a_i on $\{\bar{r}_k\}_{k=1}^N$ and $\{s_k\}_{k=1}^N$. Then $\lambda_{ij} > 0$ for all i, j.*

Proof In order to show that $\lambda_{ij} > 0$ for all i, j, I represent the market's beliefs about person i's log labor productivity as follows. Conditional on the schooling $\{s_j\}_{j=1}^N$ of person i and her siblings and the performance $\{r_j\}_{j\neq i}$ of person i's siblings but not person i, the market's beliefs about the time-invariant component $g(s_i, a_i)$ of person's i's productivity are normally distributed with mean $\mu_{ni}\big(\{s_j\}_{j=1}^N, \{r_j\}_{j\neq i}\big)$ and variance σ_{ni}^2 where:

$$\mu_{ni}\big(\{s_j\}_{j=1}^N, \{r_j\}_{j\neq i}\big) = \beta s_i + \mathbb{E}\big(a_i | \{s_j\}_{j=1}^N, \{r_j\}_{j\neq i}\big) \text{ and } \sigma_{ni}^2 = \mathbb{V}\big(a_i | \{s_j\}_{j=1}^N, \{r_j\}_{j\neq i}\big). \tag{A.87}$$

Noting that the error terms $\{\eta_{ju}\}_{u=1}^{t_j}$ in each individual's performance observations are identically distributed and independent of each other and all of the other variables in the model, the conditional expectation of a_i given $\{s_j\}_{j=1}^N$ and $\{r_j\}_{j\neq i}$ can be expressed as:

$$\mathbb{E}\big(a_i | \{s_j\}_{j=1}^N, \{r_j\}_{j\neq i}\big) = \mathbb{E}\big(a_i | \{s_j\}_{j=1}^N, \{\bar{r}_j\}_{j\neq i}\big) = \tilde{\gamma}_i + \sum_{j=1}^N \tilde{\delta}_{ij} s_j + \sum_{j\neq i} \tilde{\lambda}_{ij} \bar{r}_j, \tag{A.88}$$

where $\bar{r}_j$ denotes the sample mean of $\{r_{ju}\}_{u=1}^{t_j}$. In addition, these properties of the error terms $\{\eta_{iu}\}_{u=1}^{t_i}$ imply that the market's beliefs about $g(s_i, a_i)$ given the schooling $\{s_j\}_{j=1}^{N}$ and performance $\{r_j\}_{j=1}^{N}$ of person i and her siblings are normally distributed with mean $\mu_{bi}\big(\{s_j\}_{j=1}^{N}, \{r_j\}_{j=1}^{N}\big)$ and variance σ_{bi}^2 where:

$$\begin{aligned}&\mu_{bi}\big(\{s_j\}_{j=1}^{N}, \{r_j\}_{j=1}^{N}\big) = (1-\xi_i)\mu_{ni}\big(\{s_j\}_{j=1}^{N}, \{r_j\}_{j\neq i}\big) + \xi_i \bar{r}_i,\\ &\sigma_{bi}^2 = (\sigma_{ni}^{-2} + t_i\sigma_\eta^{-2})^{-1},\ \xi_i = t_i\sigma_\eta^{-2}\sigma_{bi}^2.\end{aligned} \tag{A.89}$$

Using equations (A.87), (A.88), and (A.89) to substitute for $\mu_{ni}\big(\{s_j\}_{j=1}^{N}, \{r_j\}_{j\neq i}\big)$ in equation (A.68), one obtains:

$$\log(v_i) = \xi_i\bar{r}_i + (1-\xi_i)\sum_{j\neq i}\tilde{\lambda}_{ij}\bar{r}_j + y_i\big(\{s_j\}_{j=1}^{N}, t_i\big), \tag{A.90}$$

where the function $y_i\big(\{s_j\}_{j=1}^{N}, t_i\big)$ is given by:

$$y_i\big(\{s_j\}_{j=1}^{N}, t_i\big) = (1-\xi_i)\left(\tilde{\gamma}_i + \beta s_i + \sum_{j=1}^{N}\tilde{\delta}_{ij}s_j\right) + \tfrac{1}{2}\sigma_{bi}^2 + h(t_i). \tag{A.91}$$

Recalling that $\hat{r}_j$ is the component of $\bar{r}_j$ orthogonal to $\{s_j\}_{j=1}^{N}$, let $\hat{r}_{-i}$ represent the vector $(\hat{r}_1, \dots, \hat{r}_{i-1}, \hat{r}_{i+1}, \dots, \hat{r}_N)'$ formed by deleting the i^{th} entry from the vector $\hat{r}$. Then the regression coefficient $\tilde{\lambda}_i = (\tilde{\lambda}_{i,1}, \dots, \tilde{\lambda}_{i,i-1}, \tilde{\lambda}_{i,i+1}, \dots, \tilde{\lambda}_{i,N})$ is given by:

$$\tilde{\lambda}_i = \mathbb{C}(a_i, \hat{r}_{-i})\mathbb{V}(\hat{r}_{-i})^{-1}. \tag{A.92}$$

Note that each of the $N-1$ elements of the vector $\mathbb{C}(a_i, \hat{r}_{-i})$ is equal to Ξ_f, which was shown to be positive in proposition A.4.5. Furthermore, the off-diagonal entries of the $(N-1)\times(N-1)$ matrix $\mathbb{V}(\hat{r}_{-i})$ are all equal to Σ_0, and the j^{th} diagonal entry of $\mathbb{V}(\hat{r}_{-i})$ is equal to Σ_j if $j < i$ and Σ_{j+1} if $j \geq i$. In proposition A.4.5, it was shown that $\Sigma_j > \Sigma_0 > 0$ for all j. Therefore, $\mathbb{V}(\hat{r}_{-i})$ has the same form as the matrix $\tilde{V}$ in proposition A.4.6; so that, the sum of the entries in each column of $\mathbb{V}(\hat{r}_{-i})^{-1}$ is positive. It follows from equation (A.92) that every element of $\tilde{\lambda}_i$ is positive, because every entry of $\mathbb{C}(a_i, \hat{r}_{-i})$ is equal to the same positive constant, and each column of $\mathbb{V}(\hat{r}_{-i})^{-1}$ has a positive sum. Thus, noting that $\xi_i \in (0,1)$ in equation (A.90), the coefficient λ_{ij}

on sibling j's average performance in sibling i's log wage is positive, because λ_{ij} is equal to ξ_i for $i = j$ and $(1 - \xi_i)\tilde{\lambda}_{ij}$ for $i \neq j$. ∎

The next result describes how the coefficient on a sibling's average performance in an individual's log wage depends on both the individual's and the sibling's ages. Suppose, for example, that sibling p is older than sibling q; so that, sibling p's average performance is a more precise indicator of an individual's underlying ability than sibling q's average performance. The result below demonstrates several intuitive facts regarding the coefficient λ_{ij} on sibling j's average performance in sibling i's log wage, which is given by equation (A.69). First, sibling p's log wage places less weight than sibling q's log wage on the average performance of any sibling $g \notin \{p, q\}$. Second, the impact of one's own average performance on one's log wage is greater for sibling p than for sibling q. Third, the coefficient on sibling q's average performance in sibling p's log wage is less than the coefficient on sibling p's average performance in sibling q's log wage. Fourth, sibling p's log wage is based more on her own average performance than on sibling q's average performance. Fifth, sibling p's average performance has a stronger influence than sibling q's average performance on the log wage of any sibling $g \notin \{p, q\}$.

Proposition A.4.8 *Suppose that learning is social. Consider two distinct siblings p and q from the same family. Let λ_{ij} denote the coefficient on sibling j's average performance $\bar{r}_j$ in the regression of sibling i's ability a_i on $\{\bar{r}_k\}_{k=1}^N$ and $\{s_k\}_{k=1}^N$. If $t_p > t_q$, then:*

1. $\lambda_{pg} < \lambda_{qg}$ *for all* $g \notin \{p, q\}$,
2. $\lambda_{qq} < \lambda_{pp}$,
3. $\lambda_{pq} < \lambda_{qp}$,
4. $\lambda_{pq} < \lambda_{pp}$,
5. $\lambda_{gq} < \lambda_{gp}$ *for all* $g \notin \{p, q\}$.

Proof I first show that $\lambda_{pg} < \lambda_{qg}$ for all $g \notin \{p, q\}$. Recall the inverse variance matrix $\mathbb{V}(\hat{r})^{-1}$ appearing in equation (A.71). From proposition A.4.5, the off-diagonal entries of $\mathbb{V}(\hat{r})$ are all equal to $\Sigma_0 > 0$, and the i^{th} diagonal entry of $\mathbb{V}(\hat{r})$ is equal to $\Sigma_i > \Sigma_0$. Let Λ_{ij} denote the $(i,j)^{\text{th}}$ entry of $\mathbb{V}(\hat{r})^{-1}$. From equation (A.79), the elements of $\mathbb{V}(\hat{r})^{-1}$ satisfy:

$$\Sigma_i \Lambda_{ij} + \Sigma_0 \sum_{k \neq i} \Lambda_{kj} = \delta_{ij}, \tag{A.93}$$

where δ_{ij} equals 1 if $i = j$ and 0 if $i \neq j$. Let e and f be two distinct indices. Subtracting the above equation for $(i,j) = (f,e)$ from the same equation for $(i,j) = (e,e)$ results in:

$$\Sigma_e \Lambda_{ee} - \Sigma_f \Lambda_{fe} + \Sigma_0(\Lambda_{fe} - \Lambda_{ee}) = 1 \quad \Leftrightarrow \quad \Lambda_{fe} = [\Lambda_{ee}(\Sigma_e - \Sigma_0) - 1]/(\Sigma_f - \Sigma_0). \tag{A.94}$$

Because $\Lambda_{fe} < 0$ from proposition A.4.6 and $(\Sigma_f - \Sigma_0) > 0$ from proposition A.4.5, it must be that $[\Lambda_{ee}(\Sigma_e - \Sigma_0) - 1] < 0$. Therefore, setting $e = g$ and $f \in \{p, q\}$ in the preceding equation, one has $\Lambda_{pg} < \Lambda_{qg}$, because $t_p > t_q$ implies that $\mathbb{V}(\hat{r}_p) < \mathbb{V}(\hat{r}_q)$ or, equivalently, $\Sigma_p < \Sigma_q$. From equation (A.71), the regression coefficient λ_{ig} is equal to:

$$\lambda_{ig} = \Xi_o \Lambda_{ig} + \Xi_f \sum_{k \neq i} \Lambda_{kg}; \tag{A.95}$$

so that, the difference $\lambda_{qg} - \lambda_{pg}$ is given by:

$$\lambda_{qg} - \lambda_{pg} = (\Xi_o - \Xi_f)(\Lambda_{qg} - \Lambda_{pg}) > 0, \tag{A.96}$$

where the inequality follows because $\Xi_o > \Xi_f$ and $\Lambda_{qg} > \Lambda_{pg}$. Thus, one has $\lambda_{pg} < \lambda_{qg}$, proving the first claim.

I now show that $\lambda_{qq} < \lambda_{pp}$. Recall from equation (A.90) that sibling i's log wage has the form:

$$\log(v_i) = \xi_i \bar{r}_i + (1 - \xi_i) \sum_{j \neq i} \tilde{\lambda}_{ij} \bar{r}_j + y_i\big(\{s_j\}_{j=1}^N, t_i\big), \tag{A.97}$$

where the coefficient $\tilde{\lambda}_i = (\tilde{\lambda}_{i,1}, \ldots, \tilde{\lambda}_{i,i-1}, \tilde{\lambda}_{i,i+1}, \ldots, \tilde{\lambda}_{i,N})$ is given as in equation (A.92) by:

$$\tilde{\lambda}_i = \mathbb{C}(a_i, \hat{r}_{-i}) \mathbb{V}(\hat{r}_{-i})^{-1}, \tag{A.98}$$

and the parameter $\xi_i \in (0,1)$ is defined as in equations (A.87) and (A.89) by:

$$\xi_i = t_i \sigma_\eta^{-2} \sigma_{bi}^2, \quad \sigma_{bi}^2 = (\sigma_{ni}^{-2} + t_i \sigma_\eta^{-2})^{-1}, \quad \sigma_{ni}^2 = \mathbb{V}\big(a_i | \{s_j\}_{j=1}^N, \{r_j\}_{j \neq i}\big). \tag{A.99}$$

Since t_p is greater than t_q, sibling p's performance history $\{r_{pu}\}_{u=1}^{t_p}$ contains more signals than sibling q's performance history $\{r_{qu}\}_{u=1}^{t_q}$. Therefore, the conditional variances must satisfy $\sigma_{np}^2 \geq \sigma_{nq}^2$, because the number of sibling q's signals on which $\mathbb{V}\big(a_p | \{s_j\}_{j=1}^N, \{r_j\}_{j \neq p}\big)$ conditions is less than the number of sibling p's signals on which $\mathbb{V}\big(a_q | \{s_j\}_{j=1}^N, \{r_j\}_{j \neq q}\big)$ conditions. It follows from $t_p \sigma_\eta^{-2} > t_q \sigma_\eta^{-2}$ and $\sigma_{np}^{-2} \leq \sigma_{nq}^{-2}$ that $\xi_p > \xi_q$ in equation (A.99). Noting that $\lambda_{ii} = \xi_i$, one obtains $\lambda_{qq} < \lambda_{pp}$, which proves the second claim.

I next show that $\lambda_{pq} < \lambda_{qp}$. The main task here is to demonstrate that the coefficient $\tilde{\lambda}_i$ defined by equation (A.98) satisfies $\tilde{\lambda}_{pq} < \tilde{\lambda}_{qp}$. Because the labeling of siblings is arbitrary, one can assume without loss of generality that $p = N - 1$ and $q = N$. From proposition A.4.5, the covariance vectors $\mathbb{C}(a_p, \hat{r}_{-p})$ and $\mathbb{C}(a_q, \hat{r}_{-q})$ and the variance matrices $\mathbb{V}(\hat{r}_{-p})$ and $\mathbb{V}(\hat{r}_{-q})$ have the following properties. Each element of $\mathbb{C}(a_p, \hat{r}_{-p})$ and $\mathbb{C}(a_q, \hat{r}_{-q})$ is equal to $\Xi_f > 0$. Moreover, all the off-diagonal entries of $\mathbb{V}(\hat{r}_{-p})$ and $\mathbb{V}(\hat{r}_{-q})$ are equal to $\Sigma_0 > 0$, and the j^{th} diagonal entries of $\mathbb{V}(\hat{r}_{-p})$ and $\mathbb{V}(\hat{r}_{-q})$ are both equal to $\Sigma_j > \Sigma_0$ if $j < N - 1$. Finally, if $j = N - 1$, then the j^{th} diagonal entry of $\mathbb{V}(\hat{r}_{-p})$ is equal to $\Sigma_q > \Sigma_0$, and the j^{th} diagonal entry of $\mathbb{V}(\hat{r}_{-q})$ is equal to $\Sigma_p > \Sigma_0$.

Let Σ_i^p and Σ_i^q denote the i^{th} diagonal entries of $\mathbb{V}(\hat{r}_{-p})$ and $\mathbb{V}(\hat{r}_{-q})$, respectively. Let Λ_{ij}^p and Λ_{ij}^q denote the $(i,j)^{\text{th}}$ entries of $\mathbb{V}(\hat{r}_{-p})^{-1}$ and $\mathbb{V}(\hat{r}_{-q})^{-1}$, respectively. Let $b \in \{p, q\}$. From equation (A.79), the elements of $\mathbb{V}(\hat{r}_{-b})^{-1}$ satisfy:

$$\Sigma_i^b \Lambda_{ij}^b + \Sigma_0 \sum_{k \neq i} \Lambda_{kj}^b = \delta_{ij}, \tag{A.100}$$

where δ_{ij} equals 1 if $i = j$ and 0 if $i \neq j$. If e and f are two distinct indices, then subtracting the above equation for $(i,j) = (f,e)$ from the same equation for $(i,j) = (e,e)$ yields:

$$\Sigma_e^b \Lambda_{ee}^b - \Sigma_f^b \Lambda_{fe}^b + \Sigma_0 (\Lambda_{fe}^b - \Lambda_{ee}^b) = 1 \quad \Leftrightarrow \quad \Lambda_{fe}^b = [\Lambda_{ee}^b (\Sigma_e^b - \Sigma_0) - 1] / (\Sigma_f^b - \Sigma_0), \tag{A.101}$$

which is analogous to equation (A.94). Using equation (A.101) with $(e,f) = (j,k)$ to substitute for Λ_{kj}^b in equation (A.100) with $j = i$, one has:

$$\Sigma_i^b \Lambda_{ii}^b + \Sigma_0 [\Lambda_{ii}^b (\Sigma_i^b - \Sigma_0) - 1] \sum_{k \neq i} (\Sigma_k^b - \Sigma_0)^{-1} = 1. \tag{A.102}$$

Solving for Λ_{ii}^b in equation (A.102) yields:

$$\Lambda_{ii}^b = \left(1 + \Sigma_0 \sum_{k \neq i} (\Sigma_k^b - \Sigma_0)^{-1}\right) \left(\Sigma_i^b + \Sigma_0 (\Sigma_i^b - \Sigma_0) \sum_{k \neq i} (\Sigma_k^b - \Sigma_0)^{-1}\right)^{-1}, \tag{A.103}$$

where $\Sigma_j^b > \Sigma_0 > 0$ for all j. If $i = N - 1$, then $\Sigma_i^p = \Sigma_q = \mathbb{V}(\hat{r}_q) > \mathbb{V}(\hat{r}_p) = \Sigma_p = \Sigma_i^q$ and $\Sigma_k^p = \Sigma_k = \Sigma_k^q$ for $k \neq i$. Therefore, equation (A.103) implies that $\Lambda_{ii}^p < \Lambda_{ii}^q$ for $i = N - 1$.

Rearranging equation (A.102), one also obtains:

$$\begin{aligned} \Sigma_0 \Lambda_{ii}^b + \Sigma_0 \sum_{k \neq i} \frac{\Lambda_{ii}^b (\Sigma_i^b - \Sigma_0) - 1}{\Sigma_k^b - \Sigma_0} &= 1 - \Lambda_{ii}^b (\Sigma_i^b - \Sigma_0) \\ \Leftrightarrow \Sigma_0 \Lambda_{ii}^b &= \Upsilon_i^b \left(1 + \Sigma_0 \sum_{k \neq i} (\Sigma_k^b - \Sigma_0)^{-1}\right), \end{aligned} \tag{A.104}$$

where Υ_i^b is defined as:

$$\Upsilon_i^b = 1 - \Lambda_{ii}^b (\Sigma_i^b - \Sigma_0). \tag{A.105}$$

If $i = N - 1$, then $\Lambda_{ii}^p < \Lambda_{ii}^q$ as shown above. Hence, equation (A.104) implies that $\Upsilon_i^p < \Upsilon_i^q$ if $i = N - 1$, noting that $\Sigma_k^b > \Sigma_0 > 0$ and $\Sigma_k^p = \Sigma_k^q$ for all $k \neq i$.

The sum of the elements in the j^{th} column of $\mathbb{V}(\hat{r}_{-b})^{-1}$ can be expressed as:

$$\sum_{k=1}^{N-1} \Lambda_{kj}^b = \Lambda_{jj}^b + [\Lambda_{jj}^b (\Sigma_j^b - \Sigma_0) - 1] \sum_{k \neq j} (\Sigma_k^b - \Sigma_0)^{-1} = \Upsilon_j^b / \Sigma_0, \tag{A.106}$$

where the first and second steps follow from equations (A.101) and (A.104), respectively. Recall that Υ_j^b is given by equation (A.105). If $j = N - 1$, then $\Upsilon_j^p < \Upsilon_j^q$ as shown above; so that, one has for $j = N - 1$:

$$\sum_{k=1}^{N-1} \Lambda_{kj}^p < \sum_{k=1}^{N-1} \Lambda_{kj}^q. \tag{A.107}$$

Therefore, using equation (A.98), the regression coefficients $\tilde{\lambda}_{pq}$ and $\tilde{\lambda}_{qp}$ are as follows for $j = N-1$:

$$\tilde{\lambda}_{pq} = \Xi_f \left(\sum_{k=1}^{N-1} \Lambda^p_{kj} \right) < \Xi_f \left(\sum_{k=1}^{N-1} \Lambda^q_{kj} \right) = \tilde{\lambda}_{qp}, \tag{A.108}$$

where the inequality is due to equation (A.107), noting that $\Xi_f > 0$ from proposition A.4.5. From equation (A.97), the coefficients λ_{pq} and λ_{qp} are given by:

$$\lambda_{pq} = (1 - \xi_p)\tilde{\lambda}_{pq} \text{ and } \lambda_{qp} = (1 - \xi_q)\tilde{\lambda}_{qp}. \tag{A.109}$$

Because it has been shown above that ξ_p, $\xi_q \in (0, 1)$ satisfy $\xi_q < \xi_p$ and that $\tilde{\lambda}_{pq}$, $\tilde{\lambda}_{qp}$ satisfy $\tilde{\lambda}_{pq} < \tilde{\lambda}_{qp}$, it follows that $\lambda_{pq} < \lambda_{qp}$, noting that $\lambda_{pq}, \lambda_{qp} > 0$ and thus $\tilde{\lambda}_{pq}, \tilde{\lambda}_{qp} > 0$ from proposition A.4.7. This completes the proof of the third claim.

I also show that $\lambda_{pq} < \lambda_{pp}$. Consider the formula for the regression coefficient λ_i in equation (A.71). The result $\lambda_{pq} < \lambda_{qp}$, which has been proved above, is equivalent to:

$$\lambda_{pq} = \Xi_o \Lambda_{pq} + \Xi_f \sum_{k \neq p} \Lambda_{kq} < \Xi_o \Lambda_{qp} + \Xi_f \sum_{k \neq q} \Lambda_{kp} = \lambda_{qp}, \tag{A.110}$$

where Λ_{ij} denotes the $(i, j)^{\text{th}}$ entry of $\mathbb{V}(\hat{r})^{-1}$. Recall that $\Xi_o > \Xi_f > 0$ from proposition A.4.5. Because $\Lambda_{pp} > 0$ and $\Lambda_{qp} < 0$ from proposition A.4.6, one obtains:

$$\lambda_{qp} = \Xi_o \Lambda_{qp} + \Xi_f \sum_{k \neq q} \Lambda_{kp} < \Xi_o \Lambda_{pp} + \Xi_f \sum_{k \neq p} \Lambda_{kp} = \lambda_{pp}. \tag{A.111}$$

Equations (A.110) and (A.111) imply that $\lambda_{pq} < \lambda_{pp}$, which proves the fourth claim.

I finally show that $\lambda_{gq} < \lambda_{gp}$ for all $g \notin \{p, q\}$. Using equation (A.90), sibling g's log wage can be expressed as:

$$\log(v_g) = \xi_g \bar{r}_g + (1 - \xi_g) \sum_{j \neq g} \tilde{\lambda}_{gj} \bar{r}_j + y_g\big(\{s_j\}_{j=1}^N, t_g\big), \tag{A.112}$$

where the coefficient $\tilde{\lambda}_g = (\tilde{\lambda}_{g,1}, \ldots, \tilde{\lambda}_{g,g-1}, \tilde{\lambda}_{g,g+1}, \ldots, \tilde{\lambda}_{g,N})$ is given as in equation (A.92) by:

$$\tilde{\lambda}_g = \mathbb{C}(a_g, \hat{r}_{-g}) \mathbb{V}(\hat{r}_{-g})^{-1}, \tag{A.113}$$

and the parameter $\xi_g \in (0,1)$ is defined in equation (A.89). In order to prove that $\lambda_{gq} < \lambda_{gp}$, I will show that $\tilde{\lambda}_{gq} < \tilde{\lambda}_{gp}$. From proposition A.4.5, every entry of $\mathbb{C}(a_g, \hat{r}_{-g})$ is equal to $\Xi_f > 0$, and every off-diagonal entry of $\mathbb{V}(\hat{r}_{-g})^{-1}$ is equal to $\Sigma_0 > 0$. Moreover, the j^{th} diagonal entry of $\mathbb{V}(\hat{r}_{-g})$ is equal to Σ_j if $j < g$ and to Σ_{j+1} if $j \geq g$, where $\Sigma_i > \Sigma_0$ for all i.

Let Σ_i^g denote the i^{th} diagonal entry of $\mathbb{V}(\hat{r}_{-g})$, and let Λ_{ij}^g denote the $(i,j)^{\text{th}}$ entry of $\mathbb{V}(\hat{r}_{-g})^{-1}$. From equation (A.79), the elements of $\mathbb{V}(\hat{r}_{-g})^{-1}$ satisfy:

$$\Sigma_i^g \Lambda_{ij}^g + \Sigma_0 \sum_{k \neq i} \Lambda_{kj}^g = \delta_{ij}, \tag{A.114}$$

where δ_{ij} equals 1 if $i = j$ and 0 if $i \neq j$. Let e and f be two distinct indices. Subtracting the above equation for $(i,j) = (f,e)$ from the same equation for $(i,j) = (e,e)$, one obtains after some rearrangement:

$$\Lambda_{fe}^g = [\Lambda_{ee}^g(\Sigma_e^g - \Sigma_0) - 1]/(\Sigma_f^g - \Sigma_0). \tag{A.115}$$

Using equation (A.115) with $(e,f) = (j,k)$ to substitute for Λ_{kj}^g in equation (A.114) with $j = i$, one obtains after solving for Λ_{ii}^g:

$$\Lambda_{ii}^g = \left(1 + \Sigma_0 \sum_{k \neq i} (\Sigma_k^g - \Sigma_0)^{-1}\right) \left(\Sigma_i^g + \Sigma_0(\Sigma_i^g - \Sigma_0) \sum_{k \neq i} (\Sigma_k^g - \Sigma_0)^{-1}\right)^{-1}. \tag{A.116}$$

Note that equations (A.115) and (A.116) are analogous to equations (A.101) and (A.103), respectively.

The sum of the elements in the j^{th} column of $\mathbb{V}(\hat{r}_{-g})^{-1}$ can be expressed as:

$$\begin{aligned} \sum_{k=1}^{N-1} \Lambda_{kj}^g &= \Lambda_{jj}^g + [\Lambda_{jj}^g(\Sigma_j^g - \Sigma_0) - 1] \sum_{k \neq j} (\Sigma_k^g - \Sigma_0)^{-1} \\ &= \left(\Sigma_j^g + \Sigma_0(\Sigma_j^g - \Sigma_0) \sum_{k \neq j} (\Sigma_k^g - \Sigma_0)^{-1}\right)^{-1}, \end{aligned} \tag{A.117}$$

where the first step uses equation (A.115) to substitute for Λ_{kj}^g if $k \neq j$, and the second step follows from substituting for Λ_{jj}^g using equation (A.116) and simplifying the resulting expression. Note that $\Sigma_i^g > \Sigma_0 > 0$ for all i in equation (A.117).

Let $\hat{p}$ and $\hat{q}$ respectively denote the columns of the matrix $\mathbb{V}(\hat{r}_{-g})$ that correspond to siblings p and q. That is, for $b \in \{p, q\}$, let $\hat{b} = b$ if $b < g$, and let $\hat{b} = b - 1$ if $b > g$. Then it follows from $t_p > t_q$ that $\Sigma^g_{\hat{p}} = \Sigma_p = \mathbb{V}(\hat{r}_p) < \mathbb{V}(\hat{r}_q) = \Sigma_q = \Sigma^g_{\hat{q}}$. Therefore, one has $\Sigma^g_{\hat{p}} < \Sigma^g_{\hat{q}}$ and $\sum_{k \neq \hat{p}} (\Sigma^g_k - \Sigma_0)^{-1} < \sum_{k \neq \hat{q}} (\Sigma^g_k - \Sigma_0)^{-1}$ in equation (A.117). Hence, the sums of the elements in the $\hat{p}^{\text{th}}$ and $\hat{q}^{\text{th}}$ columns of $\mathbb{V}(\hat{r}_{-g})^{-1}$ satisfy:

$$\sum_{k=1}^{N-1} \Lambda^g_{k\hat{q}} < \sum_{k=1}^{N-1} \Lambda^g_{k\hat{p}}. \tag{A.118}$$

From equation (A.113), the regression coefficients $\tilde{\lambda}_{gp}$ and $\tilde{\lambda}_{gq}$ are as follows:

$$\tilde{\lambda}_{gq} = \Xi_f \left(\sum_{k=1}^{N-1} \Lambda^g_{k\hat{q}} \right) < \Xi_f \left(\sum_{k=1}^{N-1} \Lambda^g_{k\hat{p}} \right) = \tilde{\lambda}_{gp}, \tag{A.119}$$

where the inequality is due to equation (A.118), noting that $\Xi_f > 0$ from proposition A.4.5. Using equation (A.112), the coefficients λ_{gp} and λ_{gq} are given by:

$$\lambda_{gp} = (1 - \xi_g)\tilde{\lambda}_{gp} \quad \text{and} \quad \lambda_{gq} = (1 - \xi_g)\tilde{\lambda}_{gq}. \tag{A.120}$$

Because $\tilde{\lambda}_{gq} < \tilde{\lambda}_{gp}$ and $\xi_g \in (0, 1)$, one has $\lambda_{gq} < \lambda_{gp}$, which proves the fifth claim. ∎

Having derived the properties of the log wage under social learning, I now turn to the problem faced by the econometrician with data on test scores, schooling, and log wages. From equation (A.69), the conditional expectation of the log wage of sibling $i \in \{p, q\}$ given z_p, z_q, and $\{s_k\}_{k=1}^N$ can be expressed as:

$$\mathbb{E}\big[\log(v_i) | z_p, z_q, \{s_k\}_{k=1}^N\big] = \sum_{j=1}^{N} \lambda_{ij} \mathbb{E}\big(a_j | z_p, z_q, \{s_k\}_{k=1}^N\big) + m_i\big(\{s_j\}_{j=1}^N, t_i\big), \tag{A.121}$$

where the function $m_i\big(\{s_j\}_{j=1}^N, t_i\big)$ is given by:

$$m_i\big(\{s_j\}_{j=1}^N, t_i\big) = \gamma_i + \beta s_i + \sum_{j=1}^{N} (\delta_{ij} + \beta\lambda_{ij}) s_j + \tfrac{1}{2}\sigma^2_{bi} + h(t_i), \tag{A.122}$$

and the parameter λ_i is defined in equation (A.71). The result below, which generalizes proposition 1.2.4, provides a testable restriction on the regression coefficients on one's own and a sibling's test scores in the conditional expectation in equation (A.121). The first part of the proposition follows directly from equation (A.71). Specifically, if sibling p is the same age as sibling q, then the parameters in equation (A.121) satisfy $\lambda_{pp} = \lambda_{qq}$, $\lambda_{pq} = \lambda_{qp}$, and $\lambda_{pj} = \lambda_{qj}$ for $j \notin \{p, q\}$; so that, the coefficients on one's own and a sibling's test scores in the conditional expectation of one's log wage given z_p, z_q, and $\{s_k\}_{k=1}^N$ are the same for both siblings p and q. The second part of the proposition indicates that if sibling p is older than sibling q, then the ratio of the coefficient on sibling q's test score to that on sibling p's test score in sibling p's log wage will typically be lower than the ratio of the coefficient on sibling p's test score to that on sibling q's test score in sibling q's log wage.

Proposition A.4.9 *Suppose that learning is social. Consider two distinct siblings p and q from the same family. Let ν_{ij} denote the regression coefficient on person j's test score in the conditional expectation of person i's log wage given z_p, z_q and $s_1, s_2, \ldots, s_N$.*

1. *If $t_p = t_q$, then $\nu_{pq}\nu_{qq} = \nu_{qp}\nu_{pp}$.*
2. *If $t_p > t_q$, then $\nu_{pq}\nu_{qq} < \nu_{qp}\nu_{pp}$.*

Proof Assume that $t_p > t_q$. For any $i \in \{p, q\}$, let $e = p$ if $i = q$, and let $e = q$ if $i = p$. From equation (A.121), the coefficients ν_{ii} and ν_{ie} in the statement of the proposition are given by:

$$\nu_{ii} = \Pi_f \lambda_{ie} + \Pi_o \lambda_{ii} + \Pi_x \sum_{k \neq i,e} \lambda_{ik} \quad \text{and} \quad \nu_{ie} = \Pi_o \lambda_{ie} + \Pi_f \lambda_{ii} + \Pi_x \sum_{k \neq i,e} \lambda_{ik}, \tag{A.123}$$

where Π_o, Π_f, and Π_x are defined in equations (A.44) and (A.59). Then the statement $\nu_{pq}\nu_{qq} < \nu_{qp}\nu_{pp}$ is equivalent to:

$$\begin{aligned} &\left(\Pi_o \lambda_{pq} + \Pi_f \lambda_{pp} + \Pi_x \sum_{k \neq p,q} \lambda_{pk}\right)\left(\Pi_f \lambda_{qp} + \Pi_o \lambda_{qq} + \Pi_x \sum_{k \neq p,q} \lambda_{qk}\right) \\ &< \left(\Pi_o \lambda_{qp} + \Pi_f \lambda_{qq} + \Pi_x \sum_{k \neq p,q} \lambda_{qk}\right)\left(\Pi_f \lambda_{pq} + \Pi_o \lambda_{pp} + \Pi_x \sum_{k \neq p,q} \lambda_{pk}\right) \end{aligned}. \tag{A.124}$$

Expanding both sides of the preceding inequality and canceling out terms appearing on both sides, one obtains after some rearrangement:

$$\begin{aligned}&\left(\Pi_f^2\lambda_{pp}\lambda_{qp}+\Pi_o^2\lambda_{pq}\lambda_{qq}\right)+\Pi_x\left(\Pi_f\Omega_a+\Pi_o\Omega_b\right)\\&<\left(\Pi_f^2\lambda_{pq}\lambda_{qq}+\Pi_o^2\lambda_{pp}\lambda_{qp}\right)+\Pi_x\left(\Pi_f\Omega_b+\Pi_o\Omega_a\right),\end{aligned}\tag{A.125}$$

where Ω_a and Ω_b are defined as:

$$\Omega_a=\lambda_{qp}\sum_{k\neq p,q}\lambda_{pk}+\lambda_{pp}\sum_{k\neq p,q}\lambda_{qk}\ \text{ and }\ \Omega_b=\lambda_{pq}\sum_{k\neq p,q}\lambda_{qk}+\lambda_{qq}\sum_{k\neq p,q}\lambda_{pk}.\tag{A.126}$$

In order to prove that $\nu_{pq}\nu_{qq}<\nu_{qp}\nu_{pp}$, one needs to show that inequality (A.125) is satisfied. Recall that $\Pi_o^2>\Pi_f^2$ from proposition A.4.2, $\lambda_{qq}>0$ and $\lambda_{pq}>0$ from proposition A.4.7, and $\lambda_{qq}<\lambda_{pp}$ and $\lambda_{pq}<\lambda_{qp}$ from proposition A.4.8. Therefore, the first term in parentheses on each side of equation (A.125) satisfies:

$$\Pi_f^2\lambda_{pp}\lambda_{qp}+\Pi_o^2\lambda_{pq}\lambda_{qq}<\Pi_f^2\lambda_{pq}\lambda_{qq}+\Pi_o^2\lambda_{pp}\lambda_{qp}.\tag{A.127}$$

From equation (A.126), the difference $\Omega_a-\Omega_b$ can be expressed as:

$$\begin{aligned}\Omega_a-\Omega_b&=\left(\lambda_{qp}-\lambda_{qq}\right)\sum_{k\neq p,q}\lambda_{pk}+\left(\lambda_{pp}-\lambda_{pq}\right)\sum_{k\neq p,q}\lambda_{qk}\\&>\left[\left(\lambda_{pp}-\lambda_{qq}\right)+\left(\lambda_{qp}-\lambda_{pq}\right)\right]\sum_{k\neq p,q}\lambda_{pk}>0.\end{aligned}\tag{A.128}$$

where the first and second inequalities follow from propositions and A.4.7 and A.4.8. In particular, the first inequality holds because $\lambda_{pp}>\lambda_{pq}$ and $\lambda_{qk}>\lambda_{pk}$ for all $k\notin\{p,q\}$, and the second inequality holds because $\lambda_{pp}>\lambda_{qq}$ and $\lambda_{qp}>\lambda_{pq}$, noting that $\lambda_{ij}>0$ for all i, j. Because $\Omega_a>\Omega_b$ as shown above and $\Pi_o>\Pi_f$ from proposition A.4.2, the second term in parentheses on each side of equation (A.125) satisfies:

$$\Pi_f\Omega_a+\Pi_o\Omega_b<\Pi_f\Omega_b+\Pi_o\Omega_a.\tag{A.129}$$

Noting that $\Pi_x>0$ from proposition A.4.3, the inequalities (A.127) and (A.129) imply that inequality (A.125) holds. Therefore, one has $\nu_{pq}\nu_{qq} < \nu_{qp}\nu_{pp}$ as desired. ∎

A.5 Data on Multiple Siblings

The analysis in the paper focuses on individuals with only one sibling, because comparatively few respondents have data on two or more interviewed siblings. Nevertheless, for those individuals with information on multiple siblings, an additional test of the social learning model can be performed by comparing the impacts of a younger and an older sibling's test scores on an individual's log wage. The setup examined here is identical to that in appendix A.4, except that the test scores of two of one's siblings are now used as regressors. That is, the assumptions about the labor market characteristics of siblings and the definitions of individual and social learning are the same as those in appendix A.4, but the empirical strategy now requires each person's log wage to be regressed on both her own and two of her siblings' test scores and schooling levels.

The first step is to characterize the regression coefficients on one's own and each sibling's test scores in the conditional expectation of one's ability given one's own test score and schooling and those of two of one's siblings. The result below extends proposition 1.2.2 to the current setting in which each individual is coupled with two instead of one sibling.

Proposition A.5.1 *Consider three distinct siblings p, q, g from the same family. Let π_a and π_b respectively denote the coefficients on own's own and each sibling's test scores in the regression of sibling i's ability a_i on z_p, z_q, z_g and s_p, s_q, s_g. Then $\pi_a > \pi_b$, $\pi_a > 0$, and $\pi_a^2 > \pi_b^2$.*

Proof Using proposition A.4.1, the component of z_i orthogonal to s_p, s_q, s_g can be expressed as:

$$\hat{z}_i = z_i - \mathbb{E}(z_i|s_i, x_i) = \{\theta_a[a_i - (\Delta_o s_i + \Delta_f x_i)] + \omega_i\} - \{\theta_a[\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})]\}, \quad \text{(A.130)}$$

where x_i denotes the average schooling of the two siblings other than sibling i included in the analysis, and the parameters Δ_o and Δ_f are defined in equation (A.45) for $N = 3$. Note that the coefficient on z_p, z_q, z_g in the regression of a_i on z_p, z_q, z_g and s_p, s_q, s_g is identical to the coefficient on $\hat{z}_p$, $\hat{z}_q$, $\hat{z}_g$ in the regression of a_i on $\hat{z}_p$, $\hat{z}_q$, $\hat{z}_g$. Therefore, the regression coefficient on the siblings' test scores in proposition A.5.1 has the form:

$$\begin{aligned}\begin{pmatrix}\pi_a & \pi_b & \pi_b\\ \pi_b & \pi_a & \pi_b\\ \pi_b & \pi_b & \pi_a\end{pmatrix} &= \mathbb{C}\left[\begin{pmatrix}a_p\\ a_q\\ a_g\end{pmatrix}, \begin{pmatrix}\hat{z}_p\\ \hat{z}_q\\ \hat{z}_g\end{pmatrix}\right]\mathbb{V}\left[\begin{pmatrix}\hat{z}_p\\ \hat{z}_q\\ \hat{z}_g\end{pmatrix}\right]^{-1}\\ &= \begin{pmatrix}\kappa_a & \kappa_b & \kappa_b\\ \kappa_b & \kappa_a & \kappa_b\\ \kappa_b & \kappa_b & \kappa_a\end{pmatrix}\begin{pmatrix}\upsilon_a & \upsilon_b & \upsilon_b\\ \upsilon_b & \upsilon_a & \upsilon_b\\ \upsilon_b & \upsilon_b & \upsilon_a\end{pmatrix}^{-1},\end{aligned} \tag{A.131}$$

where π_a and π_b respectively denote the coefficients on one's own and a sibling's test scores. In the equation above, $\mathbb{C}(a_i, \hat{z}_j)$ is represented by κ_a if $i = j$ and κ_b if $i \neq j$, and $\mathbb{C}(\hat{z}_i, \hat{z}_j)$ is represented by υ_a if $i = j$ and υ_b if $i \neq j$.

Note that the covariances κ_a and κ_b are shown to satisfy $\kappa_a > \kappa_b > 0$ in the proof of proposition A.4.2. I show here that υ_a and υ_b satisfy $\upsilon_a > \upsilon_b > 0$. For all i, the variance υ_a simplifies to:

$$\begin{aligned}\upsilon_a = \mathbb{V}(\hat{z}_i) &= \mathbb{C}\big(\{\theta_a[a_i - (\Delta_o s_i + \Delta_f x_i)] + \omega_i\} - \{\theta_a[\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})]\}, \hat{z}_i\big)\\ &= \theta_a \mathbb{C}(a_i, \hat{z}_i) + \sigma_\omega^2,\end{aligned} \tag{A.132}$$

where the first step follows from substituting for $\hat{z}_i$ using equation (A.130), and the second step uses the facts that $\hat{z}_i$ is orthogonal to s_i and x_i and that ω_i has zero covariance with a_i, s_i, and x_i. For all i, j such that $i \neq j$, the covariance υ_b reduces to:

$$\begin{aligned}\upsilon_b = \mathbb{C}(\hat{z}_i, \hat{z}_j) = \mathbb{C}\big(&\{\theta_a[a_i - (\Delta_o s_i + \Delta_f x_i)] + \omega_i\}\\ &- \{\theta_a[\mu_{ai} - (\Delta_o \mu_{si} + \Delta_f \mu_{xi})]\}, \hat{z}_j\big) = \theta_a \mathbb{C}(a_i, \hat{z}_j) + \rho_\omega \sigma_\omega^2,\end{aligned} \tag{A.133}$$

where the first step substitutes for $\hat{z}_i$ from equation (A.130), and the second step holds because $\hat{z}_j$ has zero covariance with s_i and x_i and because ω_i is uncorrelated with a_j, s_j, and x_j. In the proof of proposition A.4.2, the covariances $\mathbb{C}(a_i, \hat{z}_i)$ and $\mathbb{C}(a_i, \hat{z}_j)$ are shown to satisfy $\mathbb{C}(a_i, \hat{z}_i) > \mathbb{C}(a_i, \hat{z}_j) > 0$ for all i, j such that $i \neq j$. Therefore, it follows from equations (A.132) and (A.133) that $\upsilon_a > \upsilon_b > 0$.

Inverting the variance matrix in equation (A.131), one has:

$$\mathbb{V}\left[\begin{pmatrix} \hat{z}_p \\ \hat{z}_q \\ \hat{z}_g \end{pmatrix}\right]^{-1} = \begin{pmatrix} \iota_a & \iota_b & \iota_b \\ \iota_b & \iota_a & \iota_b \\ \iota_b & \iota_b & \iota_a \end{pmatrix}, \tag{A.134}$$

where the parameters ι_a and ι_b are given by:

$$\iota_a = (\upsilon_a + \upsilon_b)[(\upsilon_a - \upsilon_b)(\upsilon_a + 2\upsilon_b)]^{-1} \text{ and } \iota_b = -\upsilon_b/[(\upsilon_a - \upsilon_b)(\upsilon_a + 2\upsilon_b)]^{-1}. \tag{A.135}$$

Using equation (A.134) to replace the inverse variance matrix in equation (A.131), the regression coefficients π_a and π_b are given by:

$$\begin{aligned} \pi_a &= \iota_a \kappa_a + 2\iota_b \kappa_b = \frac{\kappa_a(\upsilon_a + \upsilon_b) - 2\kappa_b \upsilon_b}{(\upsilon_a - \upsilon_b)(\upsilon_a + 2\upsilon_b)}, \\ \pi_b &= \iota_b \kappa_a + \iota_a \kappa_b + \iota_b \kappa_b = \frac{\kappa_b \upsilon_a - \kappa_a \upsilon_b}{(\upsilon_a - \upsilon_b)(\upsilon_a + 2\upsilon_b)}, \end{aligned} \tag{A.136}$$

where equation (A.135) is used to substitute for ι_a and ι_b.

It is now straightforward to prove the three claims in proposition A.5.1. From equation (A.136), the difference $\pi_a - \pi_b$ is equal to:

$$\pi_a - \pi_b = (\kappa_a - \kappa_b)/(\upsilon_a - \upsilon_b), \tag{A.137}$$

which is positive because $\kappa_a > \kappa_b$ and $\upsilon_a > \upsilon_b$. Thus, $\pi_a > \pi_b$, proving the first claim. Moreover, it follows from $\kappa_a > \kappa_b > 0$ and $\upsilon_a > \upsilon_b > 0$ that both the numerator and the denominator of the expression for π_a are positive in equation (A.136). Therefore, $\pi_a > 0$, proving the second claim. Finally, using equation (A.136), the difference $\pi_a^2 - \pi_b^2$ is equal to:

$$\pi_a^2 - \pi_b^2 = \frac{(\kappa_a - \kappa_b)(\upsilon_a + 2\upsilon_b)[(\kappa_a + \kappa_b)\upsilon_a - 2\kappa_b \upsilon_b]}{(\upsilon_a - \upsilon_b)^2(\upsilon_a + 2\upsilon_b)^2}, \tag{A.138}$$

which is positive because $\kappa_a > \kappa_b > 0$ and $\upsilon_a > \upsilon_b > 0$. Hence, $\pi_a^2 > \pi_b^2$, proving the third claim. ∎

The findings here are analogous to those in proposition 1.2.2. First, one's own test score is a stronger predictor of one's ability than each sibling's test score even after controlling for one's own and both siblings' schooling. Second, the partial correlation between one's ability and one's own test score is positive given both siblings' test scores as well as one's own and both siblings' schooling. Third, the coefficient on one's own test score is larger in absolute value than the coefficient on each sibling's test score in the regression of one's ability on one's own test score and schooling and those of two of one's siblings. Note that the sign of the coefficient on each sibling's test score is indeterminate in the regression examined here.[3] As discussed in section 1.2.2, a positive correlation among siblings in the testing error ω_i can generate a negative coefficient on each sibling's test score, even though the opposite outcome might be expected because of the positive sibling correlation in ability a_i.

The second step is to analyze the problem of predicting an individual's log wage given her own and two of her siblings' test scores and schooling levels. The following result is a counterpart to propositions 1.2.3 and 1.2.4, which examine the relationship of one's log wage to one's own and a sibling's test scores under individual and social learning. In the setup here, one chooses any three siblings from a sibship with at least three members and compares the impacts of the first two siblings' test scores on the log wage of the third sibling after controlling for the third sibling's test score and the schooling of the three siblings. The first part of the result holds because an individual's ability has the same relationship to each of her siblings' test scores and schooling levels.[4] The first item in the second part follows directly from the fact that one's log wage under social learning places equal weight on the average performance of two siblings who are the same age as each other.[5]

Proposition A.5.2 *Let $N \geq 3$. Consider three distinct siblings p, q, g from the same family. Let ϖ_i denote the regression coefficient on the test score of sibling $i \in \{p, q\}$ in the conditional expectation of sibling g's log wage given z_p, z_q, z_g and s_p, s_q, s_g.*

[3] For examples in which the coefficient on each sibling's test score is respectively positive and negative, consider the limiting cases of the model in which $\rho_\omega = 0$, $\rho_\epsilon = 0$, $\rho_a \in (0, 1)$ and $\rho_a = 0$, $\rho_\epsilon = 0$, $\rho_\omega \in (0, 1)$.

[4] In particular, given three distinct siblings p, q, g from the same family, the coefficients on the test scores of siblings p and q are the same in the regression of a_g on s_p, s_q, s_g and z_p, z_q, z_g.

[5] That is, if $t_p = t_q$, then $\lambda_{gp} = \lambda_{gq}$ in equation (A.141) and so $\varpi_p - \varpi_q = 0$ in equation (A.144).

1. *If learning is individual, then* $\varpi_p = \varpi_q$.

2. *If learning is social, then:*

   (a) $\varpi_p = \varpi_q$ *for* $t_p = t_q$,

   (b) $\varpi_p > \varpi_q$ *for* $t_p > t_q$.

Proof I prove the second item in the second part of the proposition. From equations (A.69) and (A.70), sibling g's log wage under social learning can be expressed as:

$$\begin{aligned}\log(v_g) =&(\lambda_{gg}\bar{r}_g + \lambda_{gp}\bar{r}_p + \lambda_{gq}\bar{r}_q) + [(\beta + \delta_{gg})s_g + \delta_{gp}s_p + \delta_{gq}s_q] \\ &+ u\big(\{s_j\}_{j\neq p,\,q,\,g}, \{\bar{r}_j\}_{j\neq p,\,q,\,g}, t_g\big),\end{aligned} \tag{A.139}$$

where the function $u\big(\{s_j\}_{j\neq p,\,q,\,g}, \{\bar{r}_j\}_{j\neq p,\,q,\,g}, t_g\big)$ is defined as:

$$u\big(\{s_j\}_{j\neq p,\,q,\,g}, \{\bar{r}_j\}_{j\neq p,\,q,\,g}, t_g\big) = \gamma_g + \sum_{j\neq p,\,q,\,g} \lambda_{gj}\bar{r}_j + \sum_{j\neq p,\,q,\,g} \delta_{gj}s_j + \tfrac{1}{2}\sigma^2_{bg} + h(t_g). \tag{A.140}$$

Denoting $c = (s_p, s_q, s_g, z_p, z_q, z_g)$, the conditional expectation of sibling g's log wage given s_p, s_q, s_g and z_p, z_q, z_g is of the form:

$$\mathbb{E}[\log(v_g)|c] = \lambda_{gg}\mathbb{E}[a_g|c] + \lambda_{gp}\mathbb{E}[a_p|c] + \lambda_{gq}\mathbb{E}[a_q|c] + n(c), \tag{A.141}$$

where the function $n(c)$ is given by:

$$\begin{aligned}n(c) =&[\delta_{gg} + \beta(1 + \lambda_{gg})]s_g + (\delta_{gp} + \beta\lambda_{gp})s_p + (\delta_{gq} + \beta\lambda_{gq})s_q \\ &+ \mathbb{E}\big[u\big(\{s_j\}_{j\neq p,\,q,\,g}, \{\bar{r}_j\}_{j\neq p,\,q,\,g}, t_g\big)|c\big].\end{aligned} \tag{A.142}$$

Note that the regression coefficients on z_p, z_q, and z_g in the conditional expectation of $u\big(\{s_j\}_{j\neq p,\,q,\,g}$, $\{\bar{r}_j\}_{j\neq p,\,q,\,g}, t_g\big)$ given s_p, s_q, s_g and z_p, z_q, z_g are all equal to the same constant π_x. Therefore, using equation (A.141), the regression coefficients ϖ_p and ϖ_q in proposition A.5.2 can be expressed as:

$$\varpi_p = \lambda_{gp}\pi_a + (\lambda_{gg} + \lambda_{gq})\pi_b + \pi_x \;\text{ and }\; \varpi_q = \lambda_{gq}\pi_a + (\lambda_{gg} + \lambda_{gp})\pi_b + \pi_x; \tag{A.143}$$

so that, the difference $\varpi_p - \varpi_q$ is equal to:

$$\varpi_p - \varpi_q = (\pi_a - \pi_b)(\lambda_{gp} - \lambda_{gq}), \tag{A.144}$$

which is positive because $\pi_a > \pi_b$ from proposition A.5.1 and $\lambda_{gp} > \lambda_{gq}$ from proposition A.4.8 whenever $t_p > t_q$. Therefore, one has $\varpi_p > \varpi_q$ as desired. ∎

As in propositions 1.2.3 and 1.2.4, the individual and social learning models are indistinguishable from each other when the two siblings being compared are of the same age. That is, the test scores of any two siblings will each have the same impact on the log wage of a third sibling either if learning is individual and the ages of the three siblings are arbitrary or if learning is social and the ages of the first two siblings are identical. However, if one of two siblings is older than the other, then the older sibling's test score will have a greater impact than the younger sibling's test score on the log wage of a third sibling when learning is social. This prediction of the social learning model simply reflects the fact that an older sibling's average performance is a stronger predictor of one's own ability than a younger sibling's average performance because a greater number of performance observations are available on an older than on a younger sibling. Hence, if one's wage is set equal to the conditional expectation of one's labor productivity given both one's own and every sibling's schooling and performance, then an older sibling's average performance will have a higher coefficient than a younger sibling's average performance in one's log wage equation.

A.6 Schooling Coefficients in Log Wage Regression

This result below characterizes the coefficients obtained from the regression of one's log wage on one's own and a sibling's schooling. If learning is social, then the coefficients on one's own and a younger sibling's schooling in an older sibling's log wage will be the same as the corresponding coefficients on one's own and an older sibling's schooling in a younger sibling's log wage. By contrast, if learning is individual, then the outcome of this regression will fall into one of three categories, depending on the relative values of the sibling correlations in ability and in schooling. First, if the coefficient on a sibling's schooling is positive, then an older sibling will have a lower coefficient on one's own schooling than a younger sibling, and the coefficient on a younger sibling's schooling in an older sibling's log wage will be higher than vice versa. Second, if the coefficient on a sibling's schooling is zero, then an older sibling will have the same respective coefficients on one's own and a younger sibling's schooling as a younger sibling has on one's own and an older sibling's schooling. Third, if the coefficient on a sibling's schooling is negative, then an older sibling will have a higher coefficient on one's own schooling than a younger sibling, and the coefficient on a younger sibling's schooling in an older sibling's log wage will be lower than vice versa.

Proposition A.6.1 *Suppose that sibling 1 is at least as old as sibling 2 with $d \geq 0$ being the age difference between them. Let $\digamma_{ij}$ denote the regression coefficient on sibling j's schooling in the conditional expectation of sibling i's log wage given s_1 and s_2.*

1. *If learning is individual, then:*

   (a) $\digamma_{11} = \digamma_{22}$ *and* $\digamma_{12} = \digamma_{21}$ *for* $d = 0$,

   (b) $\digamma_{11} < \digamma_{22}$ *and* $\digamma_{12} > \digamma_{21} > 0$ *for* $d > 0$ *and* $\rho_a > \rho_s$,

   (c) $\digamma_{11} = \digamma_{22}$ *and* $\digamma_{12} = \digamma_{21} = 0$ *for* $d > 0$ *and* $\rho_a = \rho_s$,

   (d) $\digamma_{11} > \digamma_{22}$ *and* $\digamma_{12} < \digamma_{21} < 0$ *for* $d > 0$ *and* $\rho_a < \rho_s$.

2. *If learning is social, then* $\digamma_{11} = \digamma_{22}$ *and* $\digamma_{12} = \digamma_{21}$.

Proof I begin by proving the last three items in the first part of the proposition.[6] Using equations (1.8), (1.9), and (1.11), the log wage of sibling $i \in \{1, 2\}$ under individual learning can be expressed as:

$$\log(w_i) = \beta s_i + (1 - \chi_i)\mathbb{E}(a_i|s_i) + \chi_i(a_i + \bar{\eta}_i) + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i), \tag{A.145}$$

where χ_i and σ_{gi}^2 are defined in equation (1.9), and $\bar{\eta}_i$ denotes the sample mean of $\{\eta_{iu}\}_{u=1}^{t_i}$. Let e be the index of the sibling other than i. Calculating the conditional expectation of $\log(w_i)$ given s_i

[6]The first item in the first part of the proposition is a direct implication of the symmetric treatment of the two siblings in the model from section 1.2.

and s_e, one has:

$$\begin{aligned}\mathbb{E}[\log(w_i)|s_i,s_e] &= \beta s_i + (1-\chi_i)\mathbb{E}(a_i|s_i) + \chi_i\mathbb{E}(a_i|s_i,s_e) + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i) \\ &= \beta s_i + (1-\chi_i)\frac{\gamma\sigma_a^2}{\sigma_s^2}s_i + \chi_i\frac{\gamma\sigma_a^2}{\sigma_s^2(1-\rho_s^2)}[(1-\rho_a\rho_s)s_i + (\rho_a-\rho_s)s_e] + \varkappa_i + \tfrac{1}{2}\sigma_{gi}^2 + h(t_i)\end{aligned}, \tag{A.146}$$

where $\varkappa_i$ is a constant. Noting that $\chi_1 > \chi_2$ whenever $t_1 > t_2$, the last three items in the first part of the proposition follow from inspecting the preceding equation.

I now prove the second part of the proposition. Using equations (1.14), (1.15), and (1.17), the log wage of sibling $i \in \{1,2\}$ under social learning can be expressed as:

$$\log(v_i) = \beta s_i + \mathbb{E}(a_i|s_i,s_e,r_i,r_e) + \tfrac{1}{2}\sigma_{qi}^2 + h(t_i), \tag{A.147}$$

where σ_{qi}^2 is defined in equation (1.15). Applying the law of iterated expectations, the conditional expectation of $\log(v_i)$ given s_i and s_e is as follows:

$$\mathbb{E}[\log(v_i)|s_i,s_e] = \beta s_i + \mathbb{E}(a_i|s_i,s_e) + \tfrac{1}{2}\sigma_{qi}^2 + h(t_i). \tag{A.148}$$

The second part of the proposition is an immediate consequence of the preceding equation. ∎

A.7 Variance of Change in Log Wage Residual

The result below characterizes the variance of the change between two age levels in the residual from the regression of one's log wage on one's own and a sibling's schooling. On the one hand, if learning is individual, then the variance of the change in the log wage residual between two given ages will be the same for both an older and a younger sibling. On the other hand, if learning is social, then this quantity will be greater for an older than for a younger sibling, and the absolute difference in this quantity between an older and a younger sibling will be increasing in the size of the age gap between them.

Proposition A.7.1 *Suppose that sibling 1 is at least as old as sibling 2 with $d \geq 0$ being the age difference between them. Choose any two ages t_a and t_b such that $t_b > t_a > d$. For $i \in \{1,2\}$ and $j \in \{a,b\}$, let $y_{i,j}$ denote the log wage of sibling i at age t_j, and let $u_{i,j}$ denote the residual from* *the regression of $y_{i,j}$ on s_1 and s_2.*

1. *If learning is individual, then $\mathbb{V}(u_{1,b} - u_{1,a}) = \mathbb{V}(u_{2,b} - u_{2,a})$.*

2. *If learning is social, then:*

   (a) $\mathbb{V}(u_{1,b} - u_{1,a}) = \mathbb{V}(u_{2,b} - u_{2,a})$ *for* $d = 0$,

   (b) $\mathbb{V}(u_{1,b} - u_{1,a}) > \mathbb{V}(u_{2,b} - u_{2,a})$ *for* $d > 0$,

   (c) $\mathbb{V}(u_{1,b} - u_{1,a}) - \mathbb{V}(u_{2,b} - u_{2,a})$ *is increasing in* d*, given* t_a *and* t_b *with* $t_b > t_a > 1$.

Proof I begin by proving the second item in the second part of the proposition.[7] Let $i \in \{1, 2\}$ and $j \in \{a, b\}$. Let e be the index of the sibling other than i. The log wage residual $u_{i,j}$ is defined as:

$$u_{i,j} = y_{i,j} - \mathbb{E}(y_{i,j}|s_i, s_e); \tag{A.149}$$

so that, the variance of the difference $u_{i,b} - u_{i,a}$ in log wage residuals is equal to:

$$\mathbb{V}(u_{i,b} - u_{i,a}) = \mathbb{V}\{[y_{i,b} - y_{i,a}] - [\mathbb{E}(y_{i,b}|s_i, s_e) - \mathbb{E}(y_{i,a}|s_i, s_e)]\}. \tag{A.150}$$

Using equation (A.147) to substitute for $y_{i,a}$ and $y_{i,b}$, the first bracketed term in equation (A.150) can be expressed as:

$$\begin{aligned} y_{i,b} - y_{i,a} = \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x}) - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}) \\ + \tfrac{1}{2}[\sigma_{qi}^2(t_b, t_b + x) - \sigma_{qi}^2(t_a, t_a + x)] + [h(t_b) - h(t_a)] \end{aligned}, \tag{A.151}$$

where $\sigma_{qi}^2(t_a, t_a + x)$ and $\sigma_{qi}^2(t_b, t_b + x)$ are analogous to σ_{qi}^2 in equation (1.15), and $x = -d$ if $i = 1$, $e = 2$ and $x = d$ if $i = 2$, $e = 1$. Using equation (A.148) to substitute for $\mathbb{E}(y_{i,a}|s_i, s_e)$ and $\mathbb{E}(y_{i,b}|s_i, s_e)$, the second bracketed term in equation (A.150) can be expressed as:

$$\mathbb{E}(y_{i,b}|s_i, s_e) - \mathbb{E}(y_{i,a}|s_i, s_e) = \tfrac{1}{2}[\sigma_{qi}^2(t_b, t_b + x) - \sigma_{qi}^2(t_a, t_a + x)] + [h(t_b) - h(t_a)]. \tag{A.152}$$

[7]The first part of the proposition and the first item in the second part follow directly from the symmetric treatment of the two siblings in the model from section 1.2.

Now, the variance in equation (A.150) can be simplified as follows:

$$
\begin{aligned}
\mathbb{V}(u_{i,b} - u_{i,a}) = \ & \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x}) \\
& - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x})] \\
= \mathbb{E}\{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x}) \\
& - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}]\} \\
+ \mathbb{V}\{\mathbb{E}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x}) \\
& - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}]\} \\
= \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x}) \\
& - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}] \\
+ \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}) \\
& - \mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x})] \\
= \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}] \\
= \mathbb{E}\{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+x}] \\
& |s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}\} \\
+ \mathbb{V}\{\mathbb{E}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+x}] \\
& |s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}\} \\
= \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+x}] \\
+ \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, & \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+x})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+x}]
\end{aligned}
, \tag{A.153}
$$

where the first equality follows from noting that $\sigma_{qi}^2(t_a, t_a + x)$, $\sigma_{qi}^2(t_b, t_b + x)$ and $h(t_a)$, $h(t_b)$ are constants in equations (A.151) and (A.152), the second and fifth equalities apply the law of total variance, and the third and sixth equalities apply the law of iterated expectations.

Next, letting l be a positive integer, the conditional expectation of a_i given s_i, s_e and $\{r_{i,k}\}_{k=1}^{t_b}$, $\{r_{e,k}\}_{k=1}^{l}$ has the following form:

$$
\mathbb{E}[a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{l}] = \kappa_{o,i,l} + \kappa_{s,l}s_i + \kappa_{h,l}s_e + \kappa_{f,l}\sum_{k=1}^{t_b} r_{i,k} + \kappa_{g,l}\sum_{k=1}^{l} r_{e,k}, \tag{A.154}
$$

where all of the coefficients vary with l, and the constant depends on i as well. In addition, I define:

$$\gamma_p = \mathbb{C}(a_i, \hat{r}_{e,v}),\ \gamma_q = \mathbb{V}(\hat{r}_{e,v}),\ \gamma_r = \mathbb{C}(\hat{r}_{e,m}, \hat{r}_{e,n}), \tag{A.155}$$

where v is a positive integer, m and n are distinct positive integers, and $\hat{r}_{e,v}$, which denotes the component of $r_{e,v}$ orthogonal to s_i, s_e, and $\{r_{i,k}\}_{k=1}^{t_b}$, is given by:

$$\hat{r}_{e,v} = r_{e,v} - \mathbb{E}(r_{e,v}|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}). \tag{A.156}$$

Now, γ_q can be expressed as:

$$\begin{aligned}\gamma_q &= \mathbb{V}[(\beta s_e + a_e + \eta_{e,v}) - \mathbb{E}(r_{e,v}|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b})] \\ &= \mathbb{V}[(\beta s_e + a_e) - \mathbb{E}(r_{e,v}|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b})] + \mathbb{V}(\eta_{e,v}) > 0,\end{aligned} \tag{A.157}$$

where the second equality follows because $\eta_{e,v}$ is independent of all the other variables in the model. In addition, γ_p and γ_r can be related as follows:

$$\gamma_r = \mathbb{C}[(\beta s_e + a_e + \eta_{e,m}) - \mathbb{E}(r_{e,m}|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}), \hat{r}_{e,n}] = \mathbb{C}(a_e, \hat{r}_{e,n}) = \gamma_p, \tag{A.158}$$

where the second equality follows because $\eta_{e,m}$ is independent of all the other variables in the model and because $\hat{r}_{e,n}$ has zero covariance with s_i, s_e, and $\{r_{i,k}\}_{k=1}^{t_b}$ by definition. Moreover, the coefficient $\kappa_{g,l}$ in equation (A.154) is given by:

$$\Theta_l = \Gamma_l \Psi_l^{-1}, \tag{A.159}$$

where Θ_l is a $1 \times l$ vector each of whose entries is $\kappa_{g,l}$, Γ_l is a $1 \times l$ vector each of whose entries is γ_p, and Ψ_l is a $l \times l$ matrix each of whose diagonal entries is γ_q and each of whose off-diagonal entries is γ_r. Calculating the matrix inverse of Ψ_l, each diagonal entry of Ψ_l^{-1} is equal to:

$$\frac{\gamma_q + (l-2)\gamma_r}{(\gamma_q - \gamma_r)[\gamma_q + (l-1)\gamma_r]}, \tag{A.160}$$

and each off-diagonal entry of Ψ_l^{-1} is equal to:

$$-\frac{\gamma_r}{(\gamma_q - \gamma_r)[\gamma_q + (l-1)\gamma_r]}. \tag{A.161}$$

Hence, each element of Θ_l is equal to:

$$\begin{aligned}\kappa_{g,l} &= \gamma_p \frac{\gamma_q + (l-2)\gamma_r}{(\gamma_q - \gamma_r)[\gamma_q + (l-1)\gamma_r]} - (l-1)\gamma_p \frac{\gamma_r}{(\gamma_q - \gamma_r)[\gamma_q + (l-1)\gamma_r]} \\ &= \frac{\gamma_p}{\gamma_q + (l-1)\gamma_r}.\end{aligned} \tag{A.162}$$

Note that γ_p must be positive. Otherwise, if γ_p were non-positive, then the preceding equation would imply that $\kappa_{g,l}$ is non-positive for $l = 1$, because γ_q is positive from equation (A.157). However, this would contradict proposition A.4.7, which shows that $\kappa_{g,l}$ must be positive. Furthermore, because $\gamma_p = \gamma_r$ from equation (A.158), it must be that γ_r is also positive. Therefore, it follows from the preceding equation that $\kappa_{g,l}$ is positive and decreasing in l.

Now, consider the two variance terms following the last equality in equation (A.153). Noting that $t_a > 1$ whenever $d > 0$, let c be an integer between $-t_a + 1$ and $t_a - 2$ inclusive. First, one has:

$$\begin{aligned}&\mathbb{V}[\mathbb{E}(a_i | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+c}) | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}] \\ &= \mathbb{V}\Bigg(\kappa_{o,i,t_b+c} + \kappa_{s,t_b+c} s_i + \kappa_{h,t_b+c} s_e + \kappa_{f,t_b+c} \sum_{k=1}^{t_b} r_{i,k} \\ &\qquad + \kappa_{g,t_b+c} \sum_{k=1}^{t_b+c} r_{e,k} \Bigg| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}\Bigg) \\ &= \kappa_{g,t_b+c}^2 \mathbb{V}\Bigg(\sum_{k=t_a+c+1}^{t_b+c} r_{e,k} \Bigg| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}\Bigg) \\ &= \kappa_{g,t_b+c}^2 \mathbb{V}\Bigg(\sum_{k=t_a+c+2}^{t_b+c+1} r_{e,k} \Bigg| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}\Bigg)\end{aligned} \tag{A.163}$$

$$
\begin{aligned}
&\geq \kappa_{g,t_b+c}^2 \mathbb{V}\left(\sum_{k=t_a+c+2}^{t_b+c+1} r_{e,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\right) \\
&> \kappa_{g,t_b+c+1}^2 \mathbb{V}\left(\sum_{k=t_a+c+2}^{t_b+c+1} r_{e,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\right) \\
&= \mathbb{V}\Bigg(\kappa_{o,i,t_b+c+1} + \kappa_{s,t_b+c+1} s_i + \kappa_{h,t_b+c+1} s_e + \kappa_{f,t_b+c+1} \sum_{k=1}^{t_b} r_{i,k} \\
&\quad + \kappa_{g,t_b+c+1} \sum_{k=1}^{t_b+c+1} r_{e,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\Bigg) \\
&= \mathbb{V}[\mathbb{E}(a_i | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+c+1}) | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}]
\end{aligned}
,
$$

where the weak inequality holds because the conditional variance here cannot increase if one controls for an additional variable, and the strict inequality follows because the positive coefficient $\kappa_{g,l}$ is shown above to be decreasing in l. Second, one has:

$$
\begin{aligned}
&\mathbb{V}[\mathbb{E}(a_i | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}) | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c}] \\
&= \mathbb{V}\Bigg(\kappa_{o,i,t_a+c} + \kappa_{s,t_a+c} s_i + \kappa_{h,t_a+c} s_e + \kappa_{f,t_a+c} \sum_{k=1}^{t_b} r_{i,k} \\
&\quad + \kappa_{g,t_a+c} \sum_{k=1}^{t_a+c} r_{e,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c}\Bigg) \\
&= \kappa_{f,t_a+c}^2 \mathbb{V}\left(\sum_{k=t_a+1}^{t_b} r_{i,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c}\right) \\
&\geq \kappa_{f,t_a+c}^2 \mathbb{V}\left(\sum_{k=t_a+1}^{t_b} r_{i,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\right) \\
&\geq \kappa_{f,t_a+c+1}^2 \mathbb{V}\left(\sum_{k=t_a+1}^{t_b} r_{i,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\right) \\
&= \mathbb{V}\Bigg(\kappa_{o,i,t_a+c+1} + \kappa_{s,t_a+c+1} s_i + \kappa_{h,t_a+c+1} s_e + \kappa_{f,t_a+c+1} \sum_{k=1}^{t_b} r_{i,k} \\
&\quad + \kappa_{g,t_a+c+1} \sum_{k=1}^{t_a+c+1} r_{e,k} \middle| s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c+1}\Bigg) \\
&= \mathbb{V}[\mathbb{E}(a_i | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}) | s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c+1}]
\end{aligned}
, \tag{A.164}
$$

where the first inequality holds because the conditional variance here cannot increase if one controls for an additional variable, and the second inequality follows from the proof of proposition A.12.1, in which it is shown that the positive coefficient $\kappa_{f,l}$ is nonincreasing in l.[8]

Combining equations (A.163) and (A.164), one obtains:

$$\begin{aligned}&\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+c})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c}] \\ &\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c}] \\ &> \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+c+1})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1}] \\ &\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+c+1})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+c+1}]\end{aligned}, \tag{A.165}$$

where c is an integer between $-t_a + 1$ and $t_a - 2$ inclusive. Hence, one has:

$$\begin{aligned}&\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b-d})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d}] \\ &\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a-d}] \\ &> \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+d})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d}] \\ &\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+d}]\end{aligned}. \tag{A.166}$$

Given equation (A.153), the preceding equation implies that:

$$\mathbb{V}(u_{1,b} - u_{1,a}) > \mathbb{V}(u_{2,b} - u_{2,a}), \tag{A.167}$$

which proves the second item in the second part of the proposition.

To prove the third item in the second part of the proposition, let d_h and d_l with $d_h > d_l$ be two possible values for the age difference $d \geq 0$ between the two siblings. If $d_l = 0$, then the first and second items in the second part of the proposition imply that $\mathbb{V}(u_{1,b} - u_{1,a}) - \mathbb{V}(u_{2,b} - u_{2,a})$ is greater for $d = d_h$ than for $d = d_l$. Therefore, assume that $d_l > 0$. From equation (A.165), one

[8] Specifically, the coefficient $\xi_i \in (0, 1)$ from equation (1.15) is shown to be nonincreasing in t_e holding t_i constant, where t_i and t_e are the number of signals about one's own and a sibling's performance, respectively. Note that ξ_i is the coefficient on $\bar{r}_i$ in the conditional expectations both of $g(s_i, a_i)$ and of a_i given s_i, s_e and $\bar{r}_i$, $\bar{r}_e$.

has:

$$\begin{aligned}
&\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b-d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_h}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a-d_h}] \\
&> \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b-d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_l}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a-d_l}]
\end{aligned}, \tag{A.168}$$

and:

$$\begin{aligned}
&\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_l}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+d_l}] \\
&> \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_h}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+d_h}]
\end{aligned}. \tag{A.169}$$

Hence, one obtains:

$$\begin{aligned}
&\{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b-d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_h}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a-d_h}]\} \\
&- \{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_h}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_h})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+d_h}]\} \\
&> \{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b-d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_l}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a-d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a-d_l}]\} \\
&- \{\mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_b+d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_l}] \\
&\qquad + \mathbb{V}[\mathbb{E}(a_i|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_b}, \{r_{e,k}\}_{k=1}^{t_a+d_l})|s_i, s_e, \{r_{i,k}\}_{k=1}^{t_a}, \{r_{e,k}\}_{k=1}^{t_a+d_l}]\}
\end{aligned}. \tag{A.170}$$

Given equation (A.153), the preceding equation implies that $\mathbb{V}(u_{1,b}-u_{1,a})-\mathbb{V}(u_{2,b}-u_{2,a})$ is greater for $d = d_h$ than for $d = d_l$, which proves the third item in the second part of the proposition. ∎

A.8 Empirical Implementation of Generalized Model

This appendix provides a more general treatment of the estimation procedure in section 1.4. To simplify the exposition in the paper, each family was assumed to consist of exactly two siblings, and every sibship was assumed to enter the labor market in the same year. These two assumptions are relaxed in this appendix.[9] Consider a random sample of $I \geq 1$ sibships, each of which contains $N \geq 2$ members. The families in the sample are indexed from 1 to I, where the siblings in each family are labeled in order of decreasing age from 1 to N.[10]

Each sibship i is characterized by a year y_i, which denotes the first year that every member of the sibship is in the labor market. The random variable y_i takes values in the set $Y = \{1, 2, \ldots, D\}$. The members of sibship i are assumed to be working in each year from y_i to D. Let $t_{i,j,d}$ represent the age of sibling j from family i in year d, and let $s_{i,j}$ and $z_{i,j}$ respectively denote the schooling and the test score of sibling j from family i. The age of each individual is assumed to increase by one in each year; so that, the ages of the siblings from a given family in year zero uniquely determine their ages in years $1, 2, \ldots, D$. Letting $t_{i,0} = (t_{i,1,0}, t_{i,2,0}, \ldots, t_{i,N,0})$ represent the age structure for the N siblings from family i in year zero, the set T of possible realizations of the age structure for the siblings from a given family in year zero is taken to be finite. Every element of T is assumed to be a vector of N distinct nonnegative integers.

In addition, let $b_{i,j}$ be a $K \times 1$ vector of background variables for sibling j from family i. Although these variables were not discussed in appendix A.4, there is a simple way to formally introduce them into the analysis without changing the predictions of either learning model.[11] Assuming that $b_{i,j}$ is observable both to employers and to the econometrician, one can let the respective means $\mu_{a,i,j}$, $\mu_{\epsilon,i,j}$, $\mu_{\omega,i,j}$ of $a_{i,j}$, $\epsilon_{i,j}$, $\omega_{i,j}$ have the form:

$$\begin{aligned}(\mu_{a,i,j}, \mu_{\epsilon,i,j}, \mu_{\omega,i,j}) &= \mathbb{E}\left[(a_{i,j}, \epsilon_{i,j}, \omega_{i,j}) | \{b_{i,j}\}_{j=1}^{N}, y_i, t_{i,0}\right] \\ &= \left(\phi_{a,0} + b'_{i,j}\phi_a, \phi_{\epsilon,0} + b'_{i,j}\phi_\epsilon, \phi_{\omega,0} + b'_{i,j}\phi_\omega\right),\end{aligned} \tag{A.171}$$

[9]The discussion here is based on the presentation in appendix A.4, which extends the learning models in section 1.2 to include an arbitrary number of siblings in each family.

[10]I assume for simplicity that no two siblings from the same family have the same age.

[11]Pinkston (2009) uses a similar strategy to add demographic characteristics to a model of asymmetric employer learning.

where $\phi_{a,0}$, $\phi_{\epsilon,0}$, and $\phi_{\omega,0}$ are constants, and ϕ_a, ϕ_ϵ, and ϕ_ω are $K \times 1$ vectors of coefficients.[12]

For every family i, I construct all P_2^N distinct pairs of siblings, where P_2^N represents the number of ways of obtaining an ordered pair of distinct items from a set of N items. Each sibling pair can be represented by the triple (i, p, q), where i indexes the family from which the two siblings are drawn, and p and q are the respective labels of the first and the second siblings in the pair. For a given sibling pair (i, p, q), let $s_{i,-(p,q)}$ be the $(N-2) \times 1$ vector of schooling levels for the siblings besides p and q, and let $t_{i,-(p,q),d}$ be the $(N-2) \times 1$ vector of ages in year d for the siblings besides p and q. In addition, let $b_{i,-(p,q)}$ denote the $K(N-2) \times 1$ vector formed by stacking the $K \times 1$ vectors of background variables for each of the $N-2$ siblings other than p and q from family i.[13]

I now define the following three vectors:

$$z_{i,(p,q)} = (z_{i,p}, z_{i,q})', \quad s_{i,(p,q)} = (s_{i,p}, s_{i,q}, s'_{i,-(p,q)})', \quad b_{i,(p,q)} = (b'_{i,p}, b'_{i,q}, b'_{i,-(p,q)})', \qquad \text{(A.172)}$$

where $z_{i,(p,q)}$ consists of the test scores of siblings p and q from family i, and $s_{i,(p,q)}$ and $b_{i,(p,q)}$ respectively contain the schooling levels and background attributes of all the siblings in family i. Finally, I define two additional vectors:

$$t_{i,(p,q),d} = (t_{i,p,d}, t_{i,q,d}, t'_{i,-(p,q),d})', \quad x_{i,(p,q)} = (z'_{i,(p,q)}, s'_{i,(p,q)}, b'_{i,(p,q)})', \qquad \text{(A.173)}$$

where $t_{i,(p,q),d}$ represents the ages of all the siblings in family i in year d, and $x_{i,(p,q)}$ consists of the vectors $z_{i,(p,q)}$, $s_{i,(p,q)}$, and $b_{i,(p,q)}$. The conditional expectation of the log wage $w_{i,p,d}$ of sibling p from family i in year $d \in \{y_i, y_i + 1, \ldots, D\}$ given $x_{i,(p,q)}$ and $t_{i,(p,q),d}$ can be put in the following general form both under individual and under social learning:

$$\mathbb{E}\left(w_{i,p,d} | x_{i,(p,q)}, t_{i,(p,q),d}, y_i, t_{i,0}\right) = c(t_{i,(p,q),d}) + x'_{i,(p,q)} v(t_{i,(p,q),d}), \qquad \text{(A.174)}$$

[12] The other parameters of the model—β, σ_a^2, ρ_a, γ, σ_ϵ^2, θ_a, θ_s, σ_ω^2, ρ_ω, σ_η^2—are assumed not to vary with the realizations of $\{b_{i,j}\}_{j=1}^N$, $t_{i,0}$, and y_i. The term $h(t_{i,j,d})$ is assumed to be a function only of $t_{i,j,d}$.

[13] In each of the vectors $s_{i,-(p,q)}$, $t_{i,-(p,q),d}$, and $b_{i,-(p,q)}$, the variables of a sibling with a lower index are assumed to appear before the variables of a sibling with a higher index. For example, if $N = 5$ with $p = 4$ and $q = 2$, then $s_{i,-(p,q)} = (s_{i,1}, s_{i,3}, s_{i,5})'$, $t_{i,-(p,q),d} = (t_{i,1,d}, t_{i,3,d}, t_{i,5,d})'$, and $b_{i,-(p,q)} = (b'_{i,1}, b'_{i,3}, b'_{i,5})'$.

where $v(t_{i,(p,q),d})$ is a $[(K+1)N+2]\times 1$ coefficient vector, and $c(t_{i,(p,q),d})$ is a constant.[14] Note that both $v(t_{i,(p,q),d})$ and $c(t_{i,(p,q),d})$ can vary with the age vector $t_{i,(p,q),d}$ of the N siblings from family i in year d.[15] The preceding equation indicates that holding constant the ages of all the members of each sibship, the conditional expectation of one's log wage given one's own and a sibling's test scores as well as the schooling levels and background attributes of all the members of one's sibship is a linear function of the variables in the econometrician's information set.

I next define the two parameters of interest. For each family i, I introduce some additional variables. First, let B be the set consisting of all P_2^N ordered pairs of distinct integers between 1 and N, and let C_i be a random variable that takes on the value of each element in the set B with equal probability $1/P_2^N$.[16] Each realization of C_i designates a specific pair of siblings from family i. Second, let G_i be a random variable that takes on the value of each element of the set Y with equal probability $1/D$. Each realization of G_i represents a particular year from the set of observed dates. The random variables C_i and G_i are assumed to be independent of each other and of all the other variables in the model. Next, let H_i (resp. L_i) be a random variable that is equal to one whenever $C_i = (p,q)$ for some indices p, q with $p < q$ (resp. $p > q$) and that is equal to zero otherwise.[17] That is, H_i and L_i are indicator variables such that H_i is equal to one if and only if C_i selects a sibling pair with an older sibling listed before a younger sibling and such that L_i is equal to one if and only if C_i selects a sibling pair with a younger sibling listed before an older sibling. Finally, let W_i be a random variable that is equal to one whenever $G_i = d$ for some date d with $d \geq y_i$ and that is equal to zero otherwise. That is, W_i indicates whether all the siblings in family i have started working as of the year chosen by G_i.

For $J \in \{H, L\}$, I seek to calculate the conditional expectation of $v(t_{i,C_i,G_i})$ given that $J_i = 1$ and $W_i = 1$. Letting $\delta(\tilde{y}_i, \tilde{t}_{i,0})$ denote the proportion of sibships whose year of labor market entry is $\tilde{y}_i \in Y$ and whose age structure in year zero is $\tilde{t}_{i,0} \in T$, the expected value ν_H of $v(t_{i,C_i,G_i})$

[14]The expression for the conditional expectation in equation (A.174) is a consequence of equation (A.64) if learning is individual and of equation (A.121) if learning is social.

[15]In particular, $v(t_{i,(p,q),d})$ and $c(t_{i,(p,q),d})$ will depend on only the first element of $t_{i,(p,q),d}$ under individual learning and on all the elements of $t_{i,(p,q),d}$ under social learning.

[16]For example, if $N = 3$, then C_i has each of the following six values with equal probability $1/6$: $(1,2)$, $(2,1)$, $(1,3)$, $(3,1)$, $(2,3)$, and $(3,2)$.

[17]For example, if $N = 3$, then H_i and L_i are respectively equal to one and zero if C_i takes on the value $(1,2)$, $(1,3)$, or $(2,3)$, and H_i and L_i are respectively equal to zero and one if C_i takes on the value $(2,1)$, $(3,1)$, or $(3,2)$.

given that $H_i = 1$ and $W_i = 1$ is equal to:

$$\begin{aligned} \nu_H &= \mathbb{E}[v(t_{i,C_i,G_i})|H_i = 1, W_i = 1] \\ &= \sum_{\tilde{y}_i \in Y} \sum_{\tilde{t}_{i,0} \in T} \lambda(\tilde{y}_i, \tilde{t}_{i,0}) \sum_{d=\tilde{y}_i}^{D} \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} v(\tilde{t}_{i,(p,q),0} + d \cdot \mathbb{1}_N), \end{aligned} \tag{A.175}$$

and the expected value ν_L of $v(t_{i,C_i,G_i})$ given that $L_i = 1$ and $W_i = 1$ is equal to:

$$\begin{aligned} \nu_L &= \mathbb{E}[v(t_{i,C_i,G_i})|L_i = 1, W_i = 1] \\ &= \sum_{\tilde{y}_i \in Y} \sum_{\tilde{t}_{i,0} \in T} \lambda(\tilde{y}_i, \tilde{t}_{i,0}) \sum_{d=\tilde{y}_i}^{D} \sum_{p=2}^{N} \sum_{q=1}^{p-1} v(\tilde{t}_{i,(p,q),0} + d \cdot \mathbb{1}_N), \end{aligned} \tag{A.176}$$

where $\lambda(\tilde{y}_i, \tilde{t}_{i,0})$ is defined as:

$$\lambda(\tilde{y}_i, \tilde{t}_{i,0}) = \delta(\tilde{y}_i, \tilde{t}_{i,0}) \left(\tfrac{1}{2} P_2^N \sum_{\hat{y}_i \in Y} \sum_{\hat{t}_{i,0} \in T} \delta(\hat{y}_i, \hat{t}_{i,0})(D - \hat{y}_i + 1) \right)^{-1}, \tag{A.177}$$

and $\mathbb{1}_N$ is a $N \times 1$ vector of ones. For a randomly sampled family, ν_H (resp. ν_L) can be interpreted as the average value of the coefficient vector $v(t_{i,C_i,G_i})$ for a randomly chosen sibling pair and year, given that the first member of the sibling pair is older (resp. younger) than the second member and that all the siblings in the family have started working as of that year.[18]

It is now possible to state the following result, which is a generalization of propositions A.4.4 and A.4.9. Consider the conditional expectation function in equation (A.174) as well as the expected values of the coefficient vector in equations (A.175) and (A.176). First, if learning is individual, then the ratio of the second to the first entry of ν_H will be equal to the ratio of the second to the first entry of ν_L. That is, under individual learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will be the same as the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage. Second, if learning

[18]Observe that the first and second elements of the vector ν_H (resp. ν_L) represent the average values of the coefficients on one's own and a younger (resp. an older) sibling's test scores in the conditional expectation of an older (resp. a younger) sibling's log wage in equation (A.174).

is social, then the ratio of the second to the first entry of ν_H will be less than the ratio of the second to the first entry of ν_L, especially assuming that the first entries of ν_H and ν_L are both positive. That is, under social learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will typically be lower than the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage.

Proposition A.8.1 *Let $N \geq 2$. For $i \in \{1, 2\}$, let $\nu_{H,i}$ denote the i^{th} element of the vector ν_H in equation (A.175), and let $\nu_{L,i}$ denote the i^{th} element of the vector ν_L in equation (A.176).*

1. *If learning is individual, then $\nu_{H,2}\nu_{L,1} = \nu_{L,2}\nu_{H,1}$.*

2. *If learning is social, then $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$.*

Proof I begin by proving the first item of the proposition, which concerns the individual learning model in appendix A.4. From equation (A.64), the parameters $\nu_{H,1}$, $\nu_{H,2}$ and $\nu_{L,1}$, $\nu_{L,2}$ in the statement of the proposition have the following form under individual learning for $J \in \{H, L\}$:

$$\nu_{J,1} = \mathbb{E}[\chi(t_{i,C_i,G_i})\Pi_o | J_i = 1, W_i = 1] \text{ and } \nu_{J,2} = \mathbb{E}[\chi(t_{i,C_i,G_i})\Pi_f | J_i = 1, W_i = 1], \quad \text{(A.178)}$$

where the constants Π_o and Π_f are defined in equation (A.44). Note that $\chi(t_{i,(p,q),d})$, which only varies with the first element of $t_{i,(p,q),d}$ under individual learning, is the same as the parameter χ_i defined in equation (1.9), where its dependence on t_i was suppressed for ease of notation. Consider the identity:

$$\begin{aligned} &\{\Pi_f \mathbb{E}[\chi(t_{i,C_i,G_i})|H_i = 1, W_i = 1]\}\{\Pi_o \mathbb{E}[\chi(t_{i,C_i,G_i})|L_i = 1, W_i = 1]\} \\ &= \{\Pi_f \mathbb{E}[\chi(t_{i,C_i,G_i})|L_i = 1, W_i = 1]\}\{\Pi_o \mathbb{E}[\chi(t_{i,C_i,G_i})|H_i = 1, W_i = 1]\} \end{aligned}. \quad \text{(A.179)}$$

Noting that Π_o and Π_f are constants that do not depend on H_i, L_i, and W_i, it follows from the preceding identity that $\nu_{H,2}\nu_{L,1} = \nu_{L,2}\nu_{H,1}$ as desired, because the parameters Π_o and Π_f can be moved inside each of the expectation signs.

I next prove the second item of the proposition, which refers to the social learning model in appendix A.4. Let $C_{i,1}$ and $C_{i,2}$ be random variables such that $C_{i,1} = p$ and $C_{i,2} = q$ whenever

$C_i = (p, q)$. From equation (A.121), the parameters $\nu_{H,1}$, $\nu_{H,2}$ and $\nu_{L,1}$, $\nu_{L,2}$ in the statement of the proposition have the following form under social learning for $J \in \{H, L\}$:

$$\begin{aligned}\nu_{J,1} =& \mathbb{E}\Bigg(\Pi_f \lambda_{C_i}(t_{i,C_i,G_i}) + \Pi_o \lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i}) \\ & + \Pi_x \sum_{e \neq C_{i,1}, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i}) | J_i = 1, W_i = 1\Bigg) \\ \nu_{J,2} =& \mathbb{E}\Bigg(\Pi_o \lambda_{C_i}(t_{i,C_i,G_i}) + \Pi_f \lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i}) \\ & + \Pi_x \sum_{e \neq C_{i,1}, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i}) | J_i = 1, W_i = 1\Bigg)\end{aligned}, \tag{A.180}$$

where Π_o, Π_f, and Π_x are defined in equations (A.44) and (A.59). To be clear about the notation used here, the parameter $\lambda_{p,u}(t_{i,(p,q),d})$ in the equation above is the same as the coefficient $\lambda_{p,u}$ in proposition A.4.8, which represents the weight on sibling u's average performance in sibling p's log wage. Note that $\lambda_{p,u}(t_{i,(p,q),d})$ varies with the age vector $t_{i,(p,q),d}$ of the N siblings from family i in year d and that this dependence was suppressed in the notation of proposition A.4.8. Hence, in the current setting where the siblings in each family are all of different ages from each other and are labeled in order of decreasing age, $\lambda_{p,u}(t_{i,(p,q),d})$ denotes the coefficient on the average performance of the u^{th} oldest sibling from a given family in the log wage of the p^{th} oldest sibling in that family when the vector of ages for the siblings from that family is $t_{i,(p,q),d}$.

To shorten the notation in the remainder of the proof, let $\mathcal{E}_J[g(t_{i,C_i,G_i})]$ represent the conditional expectation $\mathbb{E}[g(t_{i,C_i,G_i})|J_i = 1, W_i = 1]$ of $g(t_{i,C_i,G_i})$ given that $J_i = 1$ and $W_i = 1$, where $J \in \{H, L\}$. Because Π_o, Π_f, and Π_x are constants that do not depend on H_i, L_i, and W_i, the parameters in equation (A.180) can be rewritten as:

$$\begin{aligned}\nu_{J,1} =&\Pi_f \mathcal{E}_J[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_o \mathcal{E}_J[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] \\ &+ \Pi_x \mathcal{E}_J\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right) \\ \nu_{J,2} =&\Pi_o \mathcal{E}_J[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_f \mathcal{E}_J[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] \\ &+ \Pi_x \mathcal{E}_J\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\end{aligned}. \tag{A.181}$$

The statement $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$ is equivalent to:

$$\begin{aligned}&\left[\Pi_o \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_f \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] + \Pi_x \mathcal{E}_H\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\right] \\ &\cdot \left[\Pi_f \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_o \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] + \Pi_x \mathcal{E}_L\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\right] \\ &< \left[\Pi_o \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_f \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] + \Pi_x \mathcal{E}_L\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\right] \\ &\cdot \left[\Pi_f \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_o \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] + \Pi_x \mathcal{E}_H\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\right]\end{aligned}. \tag{A.182}$$

Expanding both sides of the preceding inequality and canceling out terms appearing on both sides, one obtains after some rearrangement:

$$\begin{aligned}&\{\Pi_f^2 \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] \\ &+ \Pi_o^2 \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]\} + \Pi_x\{\Pi_f \widetilde{\Omega}_a + \Pi_o \widetilde{\Omega}_b\} \\ &< \{\Pi_f^2 \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] \\ &+ \Pi_o^2 \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]\} + \Pi_x\{\Pi_f \widetilde{\Omega}_b + \Pi_o \widetilde{\Omega}_a\}\end{aligned}, \tag{A.183}$$

where $\widetilde{\Omega}_a$ is defined as:

$$\begin{aligned}\widetilde{\Omega}_a = \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]\mathcal{E}_H\left(\textstyle\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right) \\ +\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_L\left(\textstyle\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\end{aligned}, \tag{A.184}$$

and $\widetilde{\Omega}_b$ is defined as:

$$\begin{aligned}\widetilde{\Omega}_b = \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]\mathcal{E}_L\left(\textstyle\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right) \\ +\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_H\left(\textstyle\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right)\end{aligned}. \tag{A.185}$$

In order to prove that $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$, I need to show that inequality (A.183) is satisfied. Consider the following three facts. First, I showed in proposition A.4.2 that $\Pi_o^2 > \Pi_f^2$. Second, I showed in proposition A.4.7 that $\lambda_{p,p}(t_{i,(p,q),d})$ and $\lambda_{p,q}(t_{i,(p,q),d})$ are positive for every family $i \in \{1, 2, \ldots, I\}$, sibling pair $(p, q) \in B$, and year $d \in \{y_i, y_i + 1, \ldots, D\}$. Therefore, one has $\mathcal{E}_J[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] > 0$ and $\mathcal{E}_J[\lambda_{C_i}(t_{i,C_i,G_i})] > 0$ for $J \in \{H, L\}$. Third, I showed in proposition A.4.8 that $\lambda_{q,q}(t_{i,(q,p),d}) < \lambda_{p,p}(t_{i,(p,q),d})$ and $\lambda_{p,q}(t_{i,(p,q),d}) < \lambda_{q,p}(t_{i,(q,p),d})$ if $t_{i,p,d} > t_{i,q,d}$. Furthermore, the random variable H_i is such that $H_i = 1$ if and only if $C_i = (p, q)$ for some p and q satisfying $t_{i,p,0} > t_{i,q,0}$, and the random variable L_i is such that $L_i = 1$ if and only if $C_i = (p, q)$ for some p and q satisfying $t_{i,p,0} < t_{i,q,0}$. Therefore, one has $\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] < \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]$ and $\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] < \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]$. It follows from these three facts that the first term in braces on each side of equation (A.183) satisfies:

$$\begin{aligned}\Pi_f^2\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_o^2\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] \\ < \Pi_f^2\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] + \Pi_o^2\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]\end{aligned}. \tag{A.186}$$

From equations (A.184) and (A.185), the difference $\widetilde{\Omega}_a - \widetilde{\Omega}_b$ can be expressed as:

$$\begin{aligned}
\widetilde{\Omega}_a - \widetilde{\Omega}_b &= \{\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] - \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\}\mathcal{E}_H\Bigg(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\Bigg) \\
&+ \{\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] - \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]\}\mathcal{E}_L\left(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\right) \\
&> \Bigg(\{\mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})] - \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]\} + \{\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] \\
&- \mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]\}\Bigg)\mathcal{E}_H\Bigg(\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})\Bigg) > 0
\end{aligned} \quad , \tag{A.187}$$

where the first and second inequalities follow from propositions A.4.7 and A.4.8. Specifically, I showed in proposition A.4.8 that $\lambda_{p,p}(t_{i,(p,q),d}) > \lambda_{p,q}(t_{i,(p,q),d})$ if $t_{i,p,d} > t_{i,q,d}$. Because $H_i = 1$ if and only if $C_i = (p,q)$ for some p and q with $t_{i,p,0} > t_{i,q,0}$, it follows that $\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] > \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]$. Moreover, I showed in proposition A.4.8 that $\lambda_{q,g}(t_{i,(q,p),d}) > \lambda_{p,g}(t_{i,(p,q),d})$ for all $g \notin \{p,q\}$ whenever $t_{i,p,d} > t_{i,q,d}$. Recall that the random variable H_i is such that $H_i = 1$ if and only if $C_i = (p,q)$ for some p and q satisfying $t_{i,p,0} > t_{i,q,0}$ and that the random variable L_i is such that $L_i = 1$ if and only if $C_i = (p,q)$ for some p and q satisfying $t_{i,p,0} < t_{i,q,0}$. Therefore, it follows that $\mathcal{E}_H[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})] < \mathcal{E}_L[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})]$. The first inequality in equation (A.187) is a consequence of the following two facts: $\mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] > \mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})]$ and $\mathcal{E}_H[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})] < \mathcal{E}_L[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})]$. In addition, I showed in proposition A.4.7 that given any $i \in \{1,2,\ldots,I\}$ and $d \in \{y_i, y_i+1,\ldots,D\}$, it must be that $\lambda_{p,e}(t_{i,(p,q),d}) > 0$ for all indices p, q, and e such that $p \neq q$ and $e \neq p,q$. Hence, one has $\mathcal{E}_H[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})] > 0$. Furthermore, I explained previously that proposition A.4.8 implies that $\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] < \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]$ and $\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] < \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]$. The second inequality in equation (A.187) results from the following three facts: $\mathcal{E}_L[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})] < \mathcal{E}_H[\lambda_{C_{i,1},C_{i,1}}(t_{i,C_i,G_i})]$, $\mathcal{E}_H[\lambda_{C_i}(t_{i,C_i,G_i})] < \mathcal{E}_L[\lambda_{C_i}(t_{i,C_i,G_i})]$, and $\mathcal{E}_H[\sum_{e \neq C_{i,1},\, C_{i,2}} \lambda_{C_{i,1},e}(t_{i,C_i,G_i})] > 0$.

Because $\widetilde{\Omega}_a > \widetilde{\Omega}_b$ from equation (A.187) and $\Pi_o > \Pi_f$ from proposition A.4.2, the second term in braces on each side of equation (A.183) satisfies:

$$\Pi_f \widetilde{\Omega}_a + \Pi_o \widetilde{\Omega}_b < \Pi_f \widetilde{\Omega}_b + \Pi_o \widetilde{\Omega}_a. \tag{A.188}$$

Noting that $\Pi_x > 0$ from proposition A.4.3, the inequalities (A.186) and (A.188) imply that inequality (A.183) holds. Therefore, one has $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$ as desired. ∎

Having shown that the main predictions of the learning models hold in aggregate, I next discuss the estimation of the conditional expectations ν_H and ν_L of the coefficient vector $v(t_{i,C_i,G_i})$. It is helpful to introduce some further notation. For a given $N \times 1$ vector t of distinct nonnegative integers, I define:

$$\begin{aligned} T_t = &\{\tilde{t}_{i,0} \in T : \exists \text{ distinct indices } \ p, q \ \text{ s.t. } \ \tilde{t}_{i,(p,q),0} + d \cdot \mathbb{1}_N = t \\ &\text{for some } \ d \in \{1, 2, \ldots, D\}.\}. \end{aligned} \tag{A.189}$$

To understand the definition of the set T_t, suppose that the observed age structure of family i in year zero is $\tilde{t}_{i,0}$. Then, for a given vector t, the observed age structure $\tilde{t}_{i,0}$ belongs to the set T_t if and only if there are two distinct siblings p and q from family i such that the associated age vector $\tilde{t}_{i,(p,q),d}$ is equal to t in some year $d \in \{1, 2, \ldots, D\}$. Note that for each element $\tilde{t}_{i,0}$ of T_t, there exists a unique pair of indices $p_t(\tilde{t}_{i,0})$, $q_t(\tilde{t}_{i,0})$ and year $d_t(\tilde{t}_{i,0})$ such that $\tilde{t}_{i,[p_t(\tilde{t}_{i,0}),q_t(\tilde{t}_{i,0})],0} + d_t(\tilde{t}_{i,0}) \cdot \mathbb{1}_N = t$. Next, for a given t, let $H_{i,t}$ (resp. $L_{i,t}$) be a random variable that is equal to one whenever $C_i = (p, q)$ and $G_i = d$ for some indices p, q and year d satisfying $p < q$ (resp. $p > q$) and $t_{i,(p,q),d} = t$ and that is equal to zero otherwise. That is, $H_{i,t}$ (resp. $L_{i,t}$) is an indicator random variable such that $H_{i,t}$ (resp. $L_{i,t}$) is equal to one if and only if C_i designates a sibling pair with an older (resp. younger) sibling listed before a younger (resp. older) sibling and the age vector associated with the sibling pair selected by C_i is equal to t in the year chosen by G_i. Now, let:

$$\begin{aligned} S_H = &\{t : \exists \ \ \tilde{t}_{i,0} \in T_t \text{ and } \tilde{y}_i \in Y \\ &\text{s.t. } p_t(\tilde{t}_{i,0}) < q_t(\tilde{t}_{i,0}), \ d_t(\tilde{t}_{i,0}) \geq \tilde{y}_i, \ \text{and } \delta(\tilde{y}_i, \tilde{t}_{i,0}) > 0\}, \end{aligned} \tag{A.190}$$

and let:

$$\begin{aligned} S_L = &\{t : \exists \ \ \tilde{t}_{i,0} \in T_t \text{ and } \tilde{y}_i \in Y \\ &\text{s.t. } p_t(\tilde{t}_{i,0}) > q_t(\tilde{t}_{i,0}), \ d_t(\tilde{t}_{i,0}) \geq \tilde{y}_i, \ \text{and } \delta(\tilde{y}_i, \tilde{t}_{i,0}) > 0\}, \end{aligned} \tag{A.191}$$

where t is assumed to be a $N \times 1$ vector of nonnegative integers in the definitions of S_H and S_L. That is, the set S_H (resp. S_L) contains the vector t if and only if there is a positive probability that both $H_{i,t} = 1$ (resp. $L_{i,t} = 1$) and $W_i = 1$.

For $J \in \{H, L\}$, let $\mu_{J,x}(t)$ represent the conditional expectation of the $[(K+1)N+2] \times 1$ random vector x_{i,C_i} given that $J_{i,t} = 1$ and $W_i = 1$, where $t \in S_J$. Then, assuming that $t \in S_J$, one can express $\mu_{J,x}(t)$ as:

$$\begin{aligned}\mu_{J,x}(t) &= \mathbb{E}(x_{i,C_i} | J_{i,t} = 1, W_i = 1) \\ &= \sum_{\tilde{t}_{i,0} \in T_t} \sum_{\tilde{y}_i = 1}^{d_t(\tilde{t}_{i,0})} \theta(\tilde{y}_i, \tilde{t}_{i,0}) \mathbb{E}(x_{i,[p_t(\tilde{t}_{i,0}), q_t(\tilde{t}_{i,0})]} | y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}),\end{aligned} \tag{A.192}$$

where $\theta(\tilde{y}_i, \tilde{t}_{i,0})$ is defined as:

$$\theta(\tilde{y}_i, \tilde{t}_{i,0}) = \delta(\tilde{y}_i, \tilde{t}_{i,0}) \left(\sum_{\hat{t}_{i,0} \in T_t} \sum_{\hat{y}_i = 1}^{d_t(\hat{t}_{i,0})} \delta(\hat{y}_i, \hat{t}_{i,0}) \right)^{-1}, \tag{A.193}$$

and $\mathbb{E}(x_{i,[p_t(\tilde{t}_{i,0}), q_t(\tilde{t}_{i,0})]} | y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0})$ is the conditional expectation of $x_{i,[p_t(\tilde{t}_{i,0}), q_t(\tilde{t}_{i,0})]}$ given that $\tilde{t}_{i,0}$ is the observed age structure of sibship i in year zero and that $\tilde{y}_i$ is the observed year of labor market entry of sibship i, where $\tilde{t}_{i,0} \in T_t$ and $\tilde{y}_i \in \{1, 2, \ldots, d_t(\tilde{t}_{i,0})\}$.

I now describe how to construct the vector of regressors used to estimate the parameters ν_H and ν_L. Fixing any integer $M \geq 0$, the set P is defined as:

$$\begin{aligned}P = &\{(e_1, e_2, \ldots, e_N) \in \mathbb{Z}^N : \\ &\sum_{c=1}^{N} e_c \leq M, \text{ where } e_c \geq 0 \text{ for every } c \in \{1, 2, \ldots, N\}.\},\end{aligned} \tag{A.194}$$

where $\mathbb{Z}$ denotes the set of integers. That is, P represents the set composed of every vector of N nonnegative integers whose entries sum to some number no greater than M. Letting $\#P$ be the number of elements in the set P, the elements of P can be labeled from 1 to $\#P$ with $e^s = (e_1^s, e_2^s, \ldots, e_N^s)$ denoting the s^{th} element of P. Then, for a given age vector $t = (t_1, t_2, \ldots, t_N)'$, let f_t denote the $\#P \times 1$ vector whose s^{th} entry is equal to the product $\prod_{c=1}^{N} t_c^{e_c^s}$; so that, the vector f_t consists of one element for every term of a M^{th}-order N-variate polynomial in t. Finally, let $h_{i,(p,q),d}$ be the $[(K+1)N + 2 + \#P] \times 1$ vector formed by stacking the vector $x_{i,(p,q)}$ on top of the vector $f_{t_{i,(p,q),d}}$. That is, I define:

$$h_{i,(p,q),d} = (x'_{i,(p,q)}, f'_{t_{i,(p,q),d}})', \tag{A.195}$$

where $x_{i,(p,q)}$ comprises the test scores of siblings p and q from family i as well as the schooling levels and background attributes of every member of sibship i, and $f_{t_{i,(p,q),d}}$ contains the terms of a multivariate polynomial in the ages in year d of all the siblings from family i.

Some further assumptions become relevant when estimating ν_H and ν_L. Let $J \in \{H, L\}$. First, each element of the conditional expectation function $\mu_{J,x}(t)$ is assumed to be adequately approximated by a M^{th}-order N-variate polynomial in the observed age vector $t \in S_J$. That is, for any $t \in S_J$, the conditional expectation $\mu_{J,x}(t)$ is specified as:

$$\mu_{J,x}(t) = \sum_{e \in P} \alpha_J^e \left(\prod_{c=1}^{N} t_c^{e_c} \right), \tag{A.196}$$

where α_J^e is a $[(K+1)N+2] \times 1$ vector that does not depend on t. Second, the matrix representing the expected value of $h_{i,C_i,G_i} h'_{i,C_i,G_i}$ given that $J_i = 1$ and $W_i = 1$ is required to be nonsingular. That is, I assume here that:

$$\text{rank}\big[\mathbb{E}(h_{i,C_i,G_i} h'_{i,C_i,G_i} | J_i = 1, W_i = 1)\big] = (K+1)N + 2 + \#P. \tag{A.197}$$

Third, the variance of x_{i,C_i} given that $J_{i,t} = 1$ and $W_i = 1$ is restricted to be a matrix of constants that do not vary with the observed age vector t. In particular, let $r_{i,(p,q),d} = x_{i,(p,q)} - \mu_{J,x}(t_{i,(p,q),d})$ if $t_{i,(p,q),d} = t$ for some $t \in S_J$. That is, given that $J_i = 1$ and $W_i = 1$, the random vector r_{i,C_i,G_i} represents the component of x_{i,C_i} orthogonal to all functions of t_{i,C_i,G_i}.[19] For any $t \in S_J$, I assume that:

$$\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_{i,t} = 1, W_i = 1) = \Sigma_{J,x}, \tag{A.198}$$

where $\Sigma_{J,x}$ is a $[(K+1)N+2] \times [(K+1)N+2]$ matrix of constants that do not depend on t.[20]

[19] To be clear, if k is a function mapping each $t \in S_J$ to a $[(K+1)N+2] \times 1$ real-valued vector $k(t)$, then the definition of $r_{i,(p,q),d}$ implies that $\mathbb{E}[r'_{i,C_i,G_i} k(t_{i,C_i,G_i}) | J_i = 1, W_i = 1] = 0$.

[20] This restriction on the conditional variance matrix can be weakened to some extent. Specifically, proposition A.8.2 remains valid if equation (A.198) is replaced by $\mathbb{E}[\Sigma_{J,x}(t_{i,C_i,G_i}) v(t_{i,C_i,G_i}) | J_i = 1, W_i = 1] = \mathbb{E}[\Sigma_{J,x}(t_{i,C_i,G_i}) | J_i =$

In addition, note that all random variables are treated as having finite first and second moments.

The following result shows that, under the assumptions above, the parameters ν_H and ν_L can be consistently estimated simply by pooling the observations on each sibling pair across every year and running ordinary least squares regressions on the resulting dataset. In particular, let:

$$\begin{aligned}\tilde{\nu}_H = &\left(V_I^{-1}\sum_{i=1}^{I}\sum_{d=y_i}^{D}\sum_{p=1}^{N-1}\sum_{q=p+1}^{N} h_{i,(p,q),d}h'_{i,(p,q),d}\right)^{-1} \\ &\left(V_I^{-1}\sum_{i=1}^{I}\sum_{d=y_i}^{D}\sum_{p=1}^{N-1}\sum_{q=p+1}^{N} h_{i,(p,q),d}w_{i,p,d}\right),\end{aligned} \tag{A.199}$$

and let:

$$\begin{aligned}\tilde{\nu}_L = &\left(V_I^{-1}\sum_{i=1}^{I}\sum_{d=y_i}^{D}\sum_{p=2}^{N}\sum_{q=1}^{p-1} h_{i,(p,q),d}h'_{i,(p,q),d}\right)^{-1} \\ &\left(V_I^{-1}\sum_{i=1}^{I}\sum_{d=y_i}^{D}\sum_{p=2}^{N}\sum_{q=1}^{p-1} h_{i,(p,q),d}w_{i,p,d}\right),\end{aligned} \tag{A.200}$$

where V_I, which represents the total number of observations, is given by:

$$V_I = \left(\tfrac{1}{2}P_2^N\sum_{i=1}^{I}(D-y_i+1)\right). \tag{A.201}$$

Let $\hat{\nu}_H$ and $\hat{\nu}_L$ be vectors containing the first $(K+1)N+2$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$, respectively. That is, $\hat{\nu}_H$ and $\hat{\nu}_L$ represent the estimated coefficients on the covariate vector $x_{i,(p,q)}$ in regressions that also control for $f_{t_{i,(p,q),d}}$. For $J \in \{H, L\}$, I show that as the number of sampled sibships I goes to infinity, the estimator $\hat{\nu}_J$ converges in probability to ν_J. The proof of proposition A.8.2 is similar to those in Wooldridge (2004).[21]

Proposition A.8.2 *Suppose that the assumptions in equations (A.196), (A.197), and (A.198) are satisfied. As the number of sampled sibships I goes to infinity, the estimators $\hat{\nu}_H$ and $\hat{\nu}_L$, which*

$1, W_i = 1]\mathbb{E}[v(t_{i,C_i,G_i})|J_i = 1, W_i = 1]$, where $\Sigma_{J,x}(t) = \mathbb{E}(r_{i,C_i,G_i}r'_{i,C_i,G_i}|J_{i,t} = 1, W_i = 1)$ for $t \in S_J$. That is, given that $J_i = 1$ and $W_i = 1$, the random coefficient vector $v(t_{i,C_i,G_i})$ is assumed to be uncorrelated with the random conditional variance matrix $\Sigma_{J,x}(t_{i,C_i,G_i})$.

[21] See especially propositions 3.2 and 6.1 in Wooldridge (2004).

consist of the first $(K+1)N+2$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$ in equations (A.199) and (A.200), respectively converge in probability to ν_H and ν_L, which are defined in equations (A.175) and (A.176).

Proof Let $C_{i,1}$ and $C_{i,2}$ be random variables such that $C_{i,1} = p$ and $C_{i,2} = q$ whenever $C_i = (p, q)$. Let $J \in \{H, L\}$. The random variable $w_{i,C_{i,1},G_i}$ can be expressed as:

$$w_{i,C_{i,1},G_i} = x'_{i,C_i}\beta_J + f'_{t_{i,C_i,G_i}}\gamma_J + e_{i,C_i,G_i}, \tag{A.202}$$

where β_J and γ_J are the unique coefficient vectors such that:

$$\begin{aligned} &\mathbb{E}(x_{i,C_i}e_{i,C_i,G_i}|J_i = 1, W_i = 1) = O_{[(K+1)N+2]\times 1}, \\ &\mathbb{E}(f_{t_{i,C_i,G_i}}e_{i,C_i,G_i}|J_i = 1, W_i = 1) = O_{\#P\times 1}, \end{aligned} \tag{A.203}$$

with $O_{[(K+1)N+2]\times 1}$ and $O_{\#P\times 1}$ being a $[(K+1)N+2]\times 1$ and a $\#P \times 1$ vector of zeros, respectively. Let $\delta_J = (\beta'_J, \gamma'_J)'$. Note that $\delta_J = [\mathbb{E}(h_{i,C_i,G_i}h'_{i,C_i,G_i}|J_i = 1, W_i = 1)]^{-1}\mathbb{E}(h_{i,C_i,G_i}w_{i,C_{i,1},G_i}| J_i = 1, W_i = 1)$ in equation (A.202). Moreover, an alternative expression for $w_{i,C_{i,1},G_i}$ is:

$$w_{i,C_{i,1},G_i} = f'_{t_{i,C_i,G_i}}\theta_J + o_{i,C_i,G_i}, \tag{A.204}$$

where θ_J is the unique coefficient vector such that:

$$\mathbb{E}(f_{t_{i,C_i,G_i}}o_{i,C_i,G_i}|J_i = 1, W_i = 1) = O_{\#P\times 1}. \tag{A.205}$$

Note that $\theta_J = [\mathbb{E}(f_{t_{i,C_i,G_i}}f'_{t_{i,C_i,G_i}}|J_i = 1, W_i = 1)]^{-1}\mathbb{E}(f_{t_{i,C_i,G_i}}w_{i,C_{i,1},G_i}|J_i = 1, W_i = 1)$ in equation (A.204). Finally, the random vector x_{i,C_i} can be decomposed as:

$$x'_{i,C_i} = f'_{t_{i,C_i,G_i}}\lambda_J + u'_{i,C_i,G_i}, \tag{A.206}$$

where λ_J is the unique $\#P \times [(K+1)N+2]$ coefficient matrix such that:

$$\mathbb{E}(f_{t_{i,C_i,G_i}}u'_{i,C_i}|J_i = 1, W_i = 1) = O_{\#P\times[(K+1)N+2]}, \tag{A.207}$$

with $O_{\#P\times[(K+1)N+2]}$ being a $\#P\times[(K+1)N+2]$ matrix of zeros. Note that $\lambda_J=[\mathbb{E}(f_{t_{i,C_i,G_i}} f'_{t_{i,C_i,G_i}}|J_i=1,W_i=1)]^{-1}\mathbb{E}(f_{t_{i,C_i,G_i}}x'_{i,C_i,G_i}|J_i=1,W_i=1)$ in equation (A.206). Because the conditional expectation function $\mu_{J,x}(t_{i,C_i,G_i})$ in equation (A.196) is assumed to be linear in the elements of $f_{t_{i,C_i,G_i}}$, one can write $\mu_{J,x}(t_{i,C_i,G_i})$ as:

$$[\mu_{J,x}(t_{i,C_i,G_i})]'=f'_{t_{i,C_i,G_i}}\lambda_J, \tag{A.208}$$

where λ_J is the same coefficient matrix in equation (A.206) as in equation (A.208).

Now, the parameter θ_J can be expressed as:

$$\begin{aligned}\theta_J&=[\mathbb{E}(f_{t_{i,C_i,G_i}}f'_{t_{i,C_i,G_i}}|J_i=1,W_i=1)]^{-1}\mathbb{E}(f_{t_{i,C_i,G_i}}w_{i,C_{i,1},G_i}|J_i=1,W_i=1)\\&=[\mathbb{E}(f_{t_{i,C_i,G_i}}f'_{t_{i,C_i,G_i}}|J_i=1,W_i=1)]^{-1}\\&\quad\mathbb{E}[f_{t_{i,C_i,G_i}}(x'_{i,C_i}\beta_J+f'_{t_{i,C_i,G_i}}\gamma_J+e_{i,C_i,G_i})|J_i=1,W_i=1]\\&=[\mathbb{E}(f_{t_{i,C_i,G_i}}f'_{t_{i,C_i,G_i}}|J_i=1,W_i=1)]^{-1}\\&\quad\mathbb{E}(f_{t_{i,C_i,G_i}}x'_{i,C_i}|J_i=1,W_i=1)\beta_J+\gamma_J=\lambda_J\beta_J+\gamma_J\end{aligned}, \tag{A.209}$$

where the second step uses equation (A.202) to substitute for $w_{i,C_{i,1},G_i}$, and the third step follows from the fact that $\mathbb{E}(f_{t_{i,C_i,G_i}}e_{i,C_i,G_i}|J_i=1,W_i=1)=O_{\#P\times 1}$. From equations (A.208) and (A.209), one has:

$$f'_{t_{i,C_i,G_i}}\theta_J=[\mu_{J,x}(t_{i,C_i,G_i})]'\beta_J+f'_{t_{i,C_i,G_i}}\gamma_J. \tag{A.210}$$

Subtracting equation (A.210) from equation (A.202) yields:

$$w_{i,C_{i,1},G_i}-f'_{t_{i,C_i,G_i}}\theta_J=[x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})]'\beta_J+e_{i,C_i,G_i}=r'_{i,C_i,G_i}\beta_J+e_{i,C_i,G_i}. \tag{A.211}$$

Multiplying the left and right sides of the preceding equation by r_{i,C_i,G_i} and taking the conditional expectation given that $J_i=1$ and $W_i=1$, one obtains:

$$\begin{aligned}&\mathbb{E}(r_{i,C_i,G_i}w_{i,C_{i,1},G_i}|J_i=1,W_i=1)-\mathbb{E}(r_{i,C_i,G_i}f'_{t_{i,C_i,G_i}}|J_i=1,W_i=1)\theta_J\\&=\mathbb{E}(r_{i,C_i,G_i}r'_{i,C_i,G_i}|J_i=1,W_i=1)\beta_J+\mathbb{E}(r_{i,C_i,G_i}e_{i,C_i,G_i}|J_i=1,W_i=1)\end{aligned}. \tag{A.212}$$

Because r_{i,C_i,G_i} is by construction orthogonal to any function of t_{i,C_i,G_i} conditional on $J_i=1$ and

$W_i = 1$, one has $\mathbb{E}(r_{i,C_i,G_i} f'_{t_{i,C_i,G_i}} | J_i = 1, W_i = 1)\theta_J = O_{[(K+1)N+2]\times 1}$, noting that $f_{t_{i,C_i,G_i}}$ is a function of t_{i,C_i,G_i}. In addition, e_{i,C_i,G_i} is orthogonal to any linear function of x_{i,C_i} and $f_{t_{i,C_i,G_i}}$ conditional on $J_i = 1$ and $W_i = 1$; so that, $\mathbb{E}(r_{i,C_i,G_i} e_{i,C_i,G_i} | J_i = 1, W_i = 1) = O_{[(K+1)N+2]\times 1}$ since r_{i,C_i,G_i} is linear in x_{i,C_i} and $f_{t_{i,C_i,G_i}}$. Therefore, equation (A.212) implies:

$$\mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1) = \mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_i = 1, W_i = 1)\beta_J; \tag{A.213}$$

so that, one has:

$$\beta_J = [\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_i = 1, W_i = 1)]^{-1} \mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1), \tag{A.214}$$

where the matrix $\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_i = 1, W_i = 1)$ is invertible because the matrix $[\mathbb{E}(h_{i,C_i,G_i} h'_{i,C_i,G_i} | J_i = 1, W_i = 1)]^{-1}$ is assumed to have full rank as in equation (A.197).

Next, I consider the vector $\mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1)$. Let $(p, q) \in B$ be any ordered pair of distinct integers between 1 and N. From equation (A.174), the log wage $w_{i,p,d}$ of sibling p from family i in year $d \in \{y_i, y_i + 1, \dots, D\}$ has the following form under both individual and social learning:

$$w_{i,p,d} = c(t_{i,(p,q),d}) + x'_{i,(p,q)} v(t_{i,(p,q),d}) + \varepsilon_{i,(p,q),d}, \tag{A.215}$$

where the error term $\varepsilon_{i,(p,q),d}$ satisfies:

$$\mathbb{E}(\varepsilon_{i,(p,q),d} | x_{i,(p,q)}, t_{i,(p,q),d}, y_i, t_{i,0}) = 0. \tag{A.216}$$

Using equation (A.215), one obtains:

$$\begin{aligned}
&\mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i{=}1, W_i{=}1) {=} \mathbb{E}\{r_{i,C_i,G_i}[c(t_{i,C_i,G_i}) \\
&+ x'_{i,C_i} v(t_{i,C_i,G_i}) {+} \varepsilon_{i,C_i,G_i}] | J_i{=}1, W_i{=}1\} \\
&{=} \mathbb{E}[r_{i,C_i,G_i} c(t_{i,C_i,G_i}) | J_i{=}1, W_i{=}1] {+} \mathbb{E}[r_{i,C_i,G_i} x'_{i,C_i} v(t_{i,C_i,G_i}) | J_i{=}1, W_i{=}1] \\
&+ \mathbb{E}[r_{i,C_i,G_i} \varepsilon_{i,C_i,G_i} | J_i{=}1, W_i{=}1]
\end{aligned}. \tag{A.217}$$

First, the conditional expectation $\mathbb{E}[r_{i,C_i,G_i}c(t_{i,C_i,G_i})|J_i=1,W_i=1]$ can be simplified as follows:

$$\begin{aligned}
&\mathbb{E}[r_{i,C_i,G_i}c(t_{i,C_i,G_i})|J_i=1,W_i=1] \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1) \\
&\qquad\quad \mathbb{E}\{[x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})]c(t_{i,C_i,G_i})|J_{i,t}=1,W_i=1\} \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1) \\
&\qquad\quad [\mathbb{E}(x_{i,C_i}|J_{i,t}=1,W_i=1)-\mu_{J,x}(t)]c(t)=O_{[(K+1)N+2]\times 1}
\end{aligned}, \tag{A.218}$$

where $O_{[(K+1)N+2]\times 1}$ is a $[(K+1)N+2]\times 1$ vector of zeros, and $\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1)$ represents the conditional probability that $t_{i,C_i,G_i}=t$ given that $J_i=1$ and $W_i=1$. In equation (A.218), the first equality follows from the law of total expectation and from replacing r_{i,C_i,G_i} with $x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})$; the second equality follows from the basic properties of the conditional expectation function; and the third equality follows from replacing $\mathbb{E}(x_{i,C_i}|J_{i,t}=1,W_i=1)$ with $\mu_{J,x}(t)$. Second, the conditional expectation $\mathbb{E}[r_{i,C_i,G_i}x'_{i,C_i}v(t_{i,C_i,G_i})|J_i=1,W_i=1]$ can be simplified as follows:

$$\begin{aligned}
&\mathbb{E}[r_{i,C_i,G_i}x'_{i,C_i}v(t_{i,C_i,G_i})|J_i=1,W_i=1] \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1) \\
&\qquad\qquad \mathbb{E}\{[x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})]x'_{i,C_i}v(t_{i,C_i,G_i})|J_{i,t}=1,W_i=1\} \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1) \\
&\qquad\qquad \Big(\mathbb{E}\{[x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})]x'_{i,C_i}|J_{i,t}=1,W_i=1\}v(t) \\
&\qquad\qquad -\mathbb{E}\{[x_{i,C_i}-\mu_{J,x}(t_{i,C_i,G_i})]\mu'_{J,x}(t_{i,C_i,G_i})|J_{i,t}=1,W_i=1\}v(t)\Big) \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1)\mathbb{E}(r_{i,C_i,G_i}r'_{i,C_i,G_i}|J_{i,t}=1,W_i=1)v(t) \\
&\quad=\sum_{t\in S_J}\Pr(t_{i,C_i,G_i}=t|J_i=1,W_i=1)\Sigma_{J,x}v(t)=\Sigma_{J,x}\mathbb{E}[v(t_{i,C_i,G_i})|J_i=1,W_i=1]
\end{aligned}. \tag{A.219}$$

In equation (A.219), the first equality follows from the law of total expectation and from substituting $x_{i,C_i} - \mu_{J,x}(t_{i,C_i,G_i})$ for r_{i,C_i,G_i}; the second equality follows from the basic properties of conditional expectations and from the fact that $\mathbb{E}\{[x_{i,C_i} - \mu_{J,x}(t_{i,C_i,G_i})]\mu'_{J,x}(t_{i,C_i,G_i})|J_{i,t} = 1, W_i = 1\}v(t) = O_{[(K+1)N+2]\times 1}$; the third equality follows from the basic properties of conditional expectations and from the definition $r_{i,C_i,G_i} = x_{i,C_i} - \mu_{J,x}(t_{i,C_i,G_i})$; the fourth equality follows from the assumption that $\mathbb{E}(r_{i,C_i,G_i}r'_{i,C_i,G_i}|J_{i,t} = 1, W_i = 1) = \Sigma_{J,x}$ in equation (A.198); and the fifth equality follows from the law of total expectation. Third, the conditional expectation $\mathbb{E}[r_{i,C_i,G_i}\varepsilon_{i,C_i,G_i}|J_i = 1, W_i = 1]$ can be simplified as follows:

$$\begin{aligned}
&\mathbb{E}[r_{i,C_i,G_i}\varepsilon_{i,C_i,G_i}|J_i = 1, W_i = 1] \\
&\quad= \sum_{\tilde{y}_i\in Y}\sum_{\tilde{t}_{i,0}\in T}\lambda(\tilde{y}_i, \tilde{t}_{i,0})\sum_{d=\tilde{y}_i}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}}\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}) \\
&\quad= \sum_{\tilde{y}_i\in Y}\sum_{\tilde{t}_{i,0}\in T}\lambda(\tilde{y}_i, \tilde{t}_{i,0})\sum_{d=\tilde{y}_i}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}}\mathbb{E}[\\
&\qquad\qquad \mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d}, y_i, t_{i,0})|y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}] \\
&\quad= \sum_{\tilde{y}_i\in Y}\sum_{\tilde{t}_{i,0}\in T}\lambda(\tilde{y}_i, \tilde{t}_{i,0})\sum_{d=\tilde{y}_i}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}}\mathbb{E}\{[x_{i,(p,q)} - \mu_{J,x}(t_{i,(p,q),d})] \\
&\qquad\qquad \mathbb{E}(\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d}, y_i, t_{i,0})|y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}\} \\
&\quad= O_{[(K+1)N+2]\times 1}
\end{aligned}, \tag{A.220}$$

where $\lambda(\tilde{y}_i, \tilde{t}_{i,0})$ is defined in equation (A.177), $p_{J,a} = 1$, $p_{J,b} = N - 1$ and $q_{J,a} = p + 1$, $q_{J,b} = N$ as in equation (A.175) if $J = H$, and $p_{J,a} = 2$, $p_{J,b} = N$ and $q_{J,a} = 1$, $q_{J,b} = p - 1$ as in equation (A.176) if $J = L$. In equation (A.220), the first and second equalities follow from the law of total expectation; the third equality follows from replacing $r_{i,(p,q),d}$ with $x_{i,(p,q)} - \mu_{J,x}(t_{i,(p,q),d})$ and from the basic properties of the conditional expectation function; and the fourth equality follows from the fact that $\mathbb{E}(\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d}, y_i, t_{i,0}) = 0$ by definition. To be clear about the notation in equation (A.220), the index (i, C_i, G_i) is treated as being random when calculating the conditional expectation $\mathbb{E}[r_{i,C_i,G_i}\varepsilon_{i,C_i,G_i}|J_i = 1, W_i = 1]$, and the index $[i, (p, q), d]$ is treated as being known when taking the conditional expectation $\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0})$. That is, one could also write $\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}) = \mathbb{E}[r_{i,C_i,G_i}\varepsilon_{i,C_i,G_i}|C_i = (p, q), G_i = d, y_i = \tilde{y}_i, t_{i,0} = \tilde{t}_{i,0}]$, where $p \neq q$ and $d \geq \tilde{y}_i$.

Substituting the results from equations (A.218), (A.219), and (A.220) into equation (A.217), one obtains:

$$\mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1) = \Sigma_{J,x} \mathbb{E}[v(t_{i,C_i,G_i}) | J_i = 1, W_i = 1]. \tag{A.221}$$

Moreover, it follows from the assumption $\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_{i,t} = 1, W_i = 1) = \Sigma_{J,x}$ in equation (A.198) that $[\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_i = 1, W_i = 1)]^{-1} = \Sigma_{J,x}^{-1}$, where $\Sigma_{J,x}$ is invertible because $\mathbb{E}(h_{i,C_i,G_i} h'_{i,C_i,G_i} | J_i = 1, W_i = 1)$ is assumed to have full rank. Therefore, the parameter β_J in equation (A.214) can be expressed as:

$$\begin{aligned} \beta_J &= [\mathbb{E}(r_{i,C_i,G_i} r'_{i,C_i,G_i} | J_i = 1, W_i = 1)]^{-1} \mathbb{E}(r_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1) \\ &= \mathbb{E}[v(t_{i,C_i,G_i}) | J_i = 1, W_i = 1] = \nu_J \end{aligned}. \tag{A.222}$$

Recall that β_J is a $[(K+1)N+2] \times 1$ vector containing the first $[(K+1)N+2]$ elements of the full coefficient vector $\delta_J = [\mathbb{E}(h_{i,C_i,G_i} h'_{i,C_i,G_i} | J_i = 1, W_i = 1)]^{-1} \mathbb{E}(h_{i,C_i,G_i} w_{i,C_{i,1},G_i} | J_i = 1, W_i = 1)$.

Now, the estimators $\tilde{\nu}_H$ and $\tilde{\nu}_L$ in equations (A.199) and (A.200) can be expressed as follows. For $\hat{y} \in Y$ and $\hat{t} \in T$, let $\chi_{i,\hat{t},\hat{y}}$ be an indicator random variable that is equal to one if $y_i = \hat{y}$ and $t_{i,0} = \hat{t}$ and that is equal to zero otherwise. Letting $J \in \{H, L\}$, one has:

$$\tilde{\nu}_J = (\tilde{\nu}_{J,1})^{-1} \tilde{\nu}_{J,2}, \tag{A.223}$$

where $\tilde{\nu}_{J,1}$ is given by:

$$\begin{aligned} \tilde{\nu}_{J,1} = (\tfrac{1}{2} P_2^N)^{-1} &\left[I^{-1} \sum_{i=1}^{I} \left(\sum_{\hat{t} \in T} \sum_{\hat{y} \in Y} \chi_{i,\hat{t},\hat{y}} \cdot (D - \hat{y} + 1) \right) \right]^{-1} \\ &\left[I^{-1} \sum_{i=1}^{I} \left(\sum_{\tilde{t} \in T} \sum_{\tilde{y} \in Y} \sum_{d=\tilde{y}}^{D} \sum_{p=p_{J,a}}^{p_{J,b}} \sum_{q=q_{J,a}}^{q_{J,b}} \chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d} h'_{i,(p,q),d} \right) \right] \end{aligned}, \tag{A.224}$$

and $\tilde{\nu}_{J,2}$ is given by:

$$\begin{aligned}\tilde{\nu}_{J,2} = (\tfrac{1}{2}P_2^N)^{-1} & \left[I^{-1} \sum_{i=1}^{I} \left(\sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \chi_{i,\hat{t},\hat{y}} \cdot (D-\hat{y}+1) \right) \right]^{-1} \\ & \left[I^{-1} \sum_{i=1}^{I} \left(\sum_{\tilde{t}\in T} \sum_{\tilde{y}\in Y} \sum_{d=\tilde{y}}^{D} \sum_{p=p_{J,a}}^{p_{J,b}} \sum_{q=q_{J,a}}^{q_{J,b}} \chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d} w_{i,p,d} \right) \right]\end{aligned}. \tag{A.225}$$

In equations (A.224) and (A.225), recall from above that $p_{J,a} = 1$, $p_{J,b} = N - 1$ and $q_{J,a} = p + 1$, $q_{J,b} = N$ as in equation (A.175) if $J = H$ and that $p_{J,a} = 2$, $p_{J,b} = N$ and $q_{J,a} = 1$, $q_{J,b} = p - 1$ as in equation (A.176) if $J = L$. Using the weak law of large numbers, one has:

$$\begin{aligned}& \text{plim}_{I\to\infty}\, I^{-1} \sum_{i=1}^{I} \left(\sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \chi_{i,\hat{t},\hat{y}} \cdot (D-\hat{y}+1) \right) = \mathbb{E} \left(\sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \chi_{i,\hat{t},\hat{y}} \cdot (D-\hat{y}+1) \right) \\ & = \sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \mathbb{E}[\chi_{i,\hat{t},\hat{y}} \cdot (D-\hat{y}+1)] = \sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \delta(\hat{y},\hat{t})(D-\hat{y}+1)\end{aligned}, \tag{A.226}$$

where $\delta(\hat{y}, \hat{t})$ denotes the probability of sampling a sibship with the observed entry date $\hat{y} \in Y$ and the observed age structure $\hat{t} \in T$. Therefore, by Slutsky's theorem, it follows that:

$$\text{plim}_{I\to\infty} \left[I^{-1} \sum_{i=1}^{I} \left(\sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \chi_{i,\hat{t},\hat{y}} \cdot (D-\hat{y}+1) \right) \right]^{-1} = \left(\sum_{\hat{t}\in T} \sum_{\hat{y}\in Y} \delta(\hat{y},\hat{t})(D-\hat{y}+1) \right)^{-1}. \tag{A.227}$$

Again, using the weak law of large numbers, one has:

$$\begin{aligned}& \text{plim}_{I\to\infty} I^{-1} \sum_{i=1}^{I} \left(\sum_{\tilde{t}\in T} \sum_{\tilde{y}\in Y} \sum_{d=\tilde{y}}^{D} \sum_{p=p_{J,a}}^{p_{J,b}} \sum_{q=q_{J,a}}^{q_{J,b}} \chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d} h'_{i,(p,q),d} \right) \\ & = \mathbb{E} \left(\sum_{\tilde{t}\in T} \sum_{\tilde{y}\in Y} \sum_{d=\tilde{y}}^{D} \sum_{p=p_{J,a}}^{p_{J,b}} \sum_{q=q_{J,a}}^{q_{J,b}} \chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d} h'_{i,(p,q),d} \right)\end{aligned} \tag{A.228}$$

$$\begin{aligned}
&= \sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \mathbb{E}(\chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d}h'_{i,(p,q),d}) \\
&= \sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \delta(\tilde{y},\tilde{t}) \cdot \mathbb{E}(h_{i,(p,q),d}h'_{i,(p,q),d}|y_i=\tilde{y},t_{i,0}=\tilde{t})
\end{aligned},$$

and, using an analogous argument, one has:

$$\begin{aligned}
&\text{plim}_{I\to\infty}\, I^{-1}\sum_{i=1}^{I}\left(\sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \chi_{i,\tilde{t},\tilde{y}} \cdot h_{i,(p,q),d}w_{i,p,d}\right) \\
&\qquad = \sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \delta(\tilde{y},\tilde{t}) \cdot \mathbb{E}(h_{i,(p,q),d}w_{i,p,d}|y_i=\tilde{y},t_{i,0}=\tilde{t})
\end{aligned}. \tag{A.229}$$

Using Slutsky's theorem, it follows from equations (A.227), (A.228), and (A.229) that:

$$\begin{aligned}
\text{plim}_{I\to\infty}\, \tilde{\nu}_{J,1} &= \sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \lambda(\tilde{y},\tilde{t}) \cdot \mathbb{E}(h_{i,(p,q),d}h'_{i,(p,q),d}|y_i=\tilde{y},t_{i,0}=\tilde{t}) \\
&= \mathbb{E}(h_{i,C_i,G_i}h'_{i,C_i,G_i}|J_i=1,W_i=1)
\end{aligned}, \tag{A.230}$$

and that:

$$\begin{aligned}
\text{plim}_{I\to\infty}\, \tilde{\nu}_{J,2} &= \sum_{\tilde{t}\in T}\sum_{\tilde{y}\in Y}\sum_{d=\tilde{y}}^{D}\sum_{p=p_{J,a}}^{p_{J,b}}\sum_{q=q_{J,a}}^{q_{J,b}} \lambda(\tilde{y},\tilde{t}) \cdot \mathbb{E}(h_{i,(p,q),d}w_{i,p,d}|y_i=\tilde{y},t_{i,0}=\tilde{t}) \\
&= \mathbb{E}(h_{i,C_i,G_i}w_{i,C_{i,1},G_i}|J_i=1,W_i=1)
\end{aligned}, \tag{A.231}$$

where $\lambda(\tilde{y},\tilde{t})$ is defined in equation (A.177). Now, by Slutsky's theorem, equations (A.230) and (A.231) along with equation (A.223) imply that:

$$\text{plim}_{I\to\infty}\, \tilde{\nu}_J = [\mathbb{E}(h_{i,C_i,G_i}h'_{i,C_i,G_i}|J_i=1,W_i=1)]^{-1}\mathbb{E}(h_{i,C_i,G_i}w_{i,C_{i,1},G_i}|J_i=1,W_i=1), \tag{A.232}$$

noting that the matrix $\mathbb{E}(h_{i,C_i,G_i}h'_{i,C_i,G_i}|J_i=1,W_i=1)$ is assumed to be nonsingular as in equation (A.197). It follows from equation (A.232) that $\text{plim}_{I\to\infty}\, \tilde{\nu}_J = \delta_J = (\beta'_J,\gamma'_J)'$, where β_J and γ_J are the regression parameters appearing in equation (A.202). In addition, recall from equation (A.222) that $\beta_J = \nu_J$. Therefore, as desired, the first $[(K+1)N+2]$ elements of $\tilde{\nu}_J$ converge in probability to ν_J. $\blacksquare$

A.9 Measurement Error in Schooling

This appendix describes how classical measurement error in the schooling variable affects the outcome of the test proposed in section 1.2.2 for determining whether ability or schooling is more highly correlated among siblings. As in section 1.2.1, true schooling and test scores are given by equations (1.3) and (1.5), respectively. Each sibling's reported schooling e_i, which can now differ from her true schooling s_i, is specified as:

$$e_i = s_i + \upsilon_i, \tag{A.233}$$

where υ_i, which has variance $\sigma_\upsilon^2 \geq 0$, is uncorrelated with (a_1, a_2), (s_1, s_2), and (z_1, z_2).

The following result, which extends proposition 1.2.1, characterizes the coefficient obtained from the regression of one's test score on one's own and a sibling's reported schooling. The formula below indicates that if schooling is measured with error, then the test in section 1.2.2 can be biased towards either a negative or a positive outcome relative to the result when schooling is not measured with error, depending on whether the measurement error in schooling is more or less correlated than true schooling. However, the test remains valid if the measurement error in schooling has the same correlation as true schooling.

Proposition A.9.1 *The regression coefficient of $(z_1, z_2)'$ on $(e_1, e_2)'$ is given by:*

$$\begin{aligned} &\begin{pmatrix} \tau_o & \tau_f \\ \tau_f & \tau_o \end{pmatrix} \\ &= \frac{1}{\sigma_e^4(1-\rho_e^2)} \left[\theta_s \sigma_s^2 \sigma_e^2 \begin{pmatrix} 1-\rho_s\rho_e & \rho_s - \rho_e \\ \rho_s - \rho_e & 1-\rho_s\rho_e \end{pmatrix} + \theta_a \gamma \sigma_a^2 \sigma_e^2 \begin{pmatrix} 1-\rho_a\rho_e & \rho_a - \rho_e \\ \rho_a - \rho_e & 1-\rho_a\rho_e \end{pmatrix} \right]. \end{aligned} \tag{A.234}$$

Proof The conditional expectation of $(z_1, z_2)'$ given $(e_1, e_2)'$ is:

$$\begin{aligned}&\mathbb{E}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mid \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right] \\ &= \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \end{pmatrix} + \mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} - \begin{pmatrix} \mu_{e1} \\ \mu_{e2} \end{pmatrix}\right],\end{aligned} \tag{A.235}$$

where the regression coefficient is given by:

$$\begin{aligned}&\mathbb{C}\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}\right]^{-1} \\ &= \left[\theta_s \sigma_s^2 \begin{pmatrix} 1 & \rho_s \\ \rho_s & 1 \end{pmatrix} + \theta_a \gamma \sigma_a^2 \begin{pmatrix} 1 & \rho_a \\ \rho_a & 1 \end{pmatrix}\right] \begin{pmatrix} \sigma_e^2 & \rho_e \sigma_e^2 \\ \rho_e \sigma_e^2 & \sigma_e^2 \end{pmatrix}^{-1}.\end{aligned} \tag{A.236}$$

Inverting the variance matrix and rearranging terms leads to the formula in equation (A.234). ∎

A.10 Asymmetries in Variance of Testing Error

This appendix examines how a difference between siblings in the variances of their testing errors affects the coefficients on one's own and a sibling's test scores when a person's log wage is regressed on both her own and her sibling's test scores and schooling levels. The statistical relationships among the characteristics of any pair of siblings are as specified in section 1.2.1, except that the variance of the testing error ω_i in equation (1.5) can now differ between the two siblings. In particular, I let $\sigma_{\omega_i}^2 > 0$ denote the variance of the testing error ω_i for sibling $i \in \{1, 2\}$. As in section 1.2.3, I focus on the case where employer learning is individual. In addition, to simplify the comparison between the log wages of the two siblings, I consider the log wage of each sibling upon reaching a given age level t, which is similar to the procedure used in appendix A.2. In this case, as stated in the first item of proposition A.2.1, it follows directly from the symmetric treatment of the two siblings that there will be no differences between them in the coefficients obtained from the regression of each sibling's log wage at a given age on her own and her sibling's schooling and test scores. However, if, for example, sibling 1 has a higher variance in the testing error than sibling 2, then the proposition below shows that in this regression, sibling 1 will have a higher coefficient on her sibling's test score and a lower coefficient on her own test score than sibling 2. A symmetric result holds if $\sigma^2_{\omega_1} < \sigma^2_{\omega_2}$ instead of $\sigma^2_{\omega_1} > \sigma^2_{\omega_2}$.

Proposition A.10.1 *Suppose that learning is individual. Let ϑ_{ij} denote the regression coefficient on sibling j's test score in the conditional expectation of sibling i's log wage at age t given s_1, s_2 and z_1, z_2.*

1. *If $\sigma^2_{\omega_1} = \sigma^2_{\omega_2}$, then $\vartheta_{12} = \vartheta_{21}$ and $\vartheta_{11} = \vartheta_{22}$.*
2. *If $\sigma^2_{\omega_1} > \sigma^2_{\omega_2}$, then $\vartheta_{12} > \vartheta_{21}$ and $\vartheta_{11} < \vartheta_{22}$.*

Proof From equation (1.12), the conditional expectation of sibling i's log wage at age t given s_1, s_2 and z_1, z_2 has the following form under individual learning:

$$\mathbb{E}[\log(w_i)|s_1, s_2, z_1, z_2] = \chi\mathbb{E}(a_i|s_1, s_2, z_1, z_2) + f_i(s_i, t), \tag{A.237}$$

where $\chi > 0$ is given by χ_i in equation (1.9) for $t_i = t$, and $f_i(s_i, t)$ is defined in equation (1.13). Therefore, it simply remains to characterize the conditional expectation of a_i given s_1, s_2 and z_1, z_2. Letting $\hat{z}_i$ denote the component of z_i orthogonal to s_1 and s_2, the coefficients on $\hat{z}_1$ and $\hat{z}_2$ in a regression on $\hat{z}_1$, $\hat{z}_2$ are the same as the coefficients on z_1 and z_2 in a regression on s_1, s_2 and z_1, z_2. Hence, the regression coefficients in the statement of the proposition can be expressed as:

$$\begin{pmatrix} \vartheta_{11} & \vartheta_{12} \\ \vartheta_{21} & \vartheta_{22} \end{pmatrix} = \mathbb{C}\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right]^{-1}, \tag{A.238}$$

with the inverse variance matrix having the form:

$$\mathbb{V}\left[\begin{pmatrix} \hat{z}_1 \\ \hat{z}_2 \end{pmatrix}\right]^{-1} = [\sigma^2_{\hat{z}_1}\sigma^2_{\hat{z}_2}(1-\rho^2_{\hat{z}})]^{-1} \begin{pmatrix} \sigma^2_{\hat{z}_2} & -\rho_{\hat{z}}\sigma_{\hat{z}_1}\sigma_{\hat{z}_2} \\ -\rho_{\hat{z}}\sigma_{\hat{z}_1}\sigma_{\hat{z}_2} & \sigma^2_{\hat{z}_1} \end{pmatrix}, \tag{A.239}$$

where $\sigma^2_{\hat{z}_1}$ and $\sigma^2_{\hat{z}_2}$ denote the respective variances of $\hat{z}_1$ and $\hat{z}_2$, and $\rho_{\hat{z}}$ represents the correlation coefficient between $\hat{z}_1$ and $\hat{z}_2$. Denoting $\hat{\tau} = [\sigma^2_{\hat{z}_1}\sigma^2_{\hat{z}_2}(1-\rho^2_{\hat{z}})]^{-1} > 0$, let $\kappa_c = \hat{\tau}\mathbb{C}(a_1, \hat{z}_1) =$

$\hat{\tau}\mathbb{C}(a_2, \hat{z}_2)$ and $\kappa_d = \hat{\tau}\mathbb{C}(a_1, \hat{z}_2) = \hat{\tau}\mathbb{C}(a_2, \hat{z}_1)$. From equations (A.238) and (A.239), the parameters ϑ_{12} and ϑ_{21} are given by:

$$\vartheta_{12} = \kappa_d \sigma^2_{\hat{z}_1} - \kappa_c \rho_{\hat{z}} \sigma_{\hat{z}_1} \sigma_{\hat{z}_2} \text{ and } \vartheta_{21} = \kappa_d \sigma^2_{\hat{z}_2} - \kappa_c \rho_{\hat{z}} \sigma_{\hat{z}_1} \sigma_{\hat{z}_2}, \tag{A.240}$$

and the parameters ϑ_{11} and ϑ_{22} are given by:

$$\vartheta_{11} = \kappa_c \sigma^2_{\hat{z}_2} - \kappa_d \rho_{\hat{z}} \sigma_{\hat{z}_1} \sigma_{\hat{z}_2} \text{ and } \vartheta_{22} = \kappa_c \sigma^2_{\hat{z}_1} - \kappa_d \rho_{\hat{z}} \sigma_{\hat{z}_1} \sigma_{\hat{z}_2}. \tag{A.241}$$

Note that the covariances $\mathbb{C}(a_1, \hat{z}_1) = \mathbb{C}(a_2, \hat{z}_2)$ and $\mathbb{C}(a_1, \hat{z}_2) = \mathbb{C}(a_2, \hat{z}_1)$ in equation (A.238) are the same as those appearing in the proof of proposition 1.2.2, where they are shown to satisfy $\mathbb{C}(a_i, \hat{z}_i) > 0$ and $\mathbb{C}(a_i, \hat{z}_j) > 0$ for all i, j such that $i \neq j$. Therefore, $\kappa_c > 0$ and $\kappa_d > 0$ in equations (A.240) and (A.241). Moreover, $\sigma^2_{\hat{z}_1} = \sigma^2_{\hat{z}_2}$ whenever $\sigma^2_{\omega_1} = \sigma^2_{\omega_2}$, and $\sigma^2_{\hat{z}_1} > \sigma^2_{\hat{z}_2}$ whenever $\sigma^2_{\omega_1} > \sigma^2_{\omega_2}$. Thus, it follows from equations (A.240) and (A.241) that $\vartheta_{12} = \vartheta_{21}$ and $\vartheta_{11} = \vartheta_{22}$ if $\sigma^2_{\omega_1} = \sigma^2_{\omega_2}$ and that $\vartheta_{12} > \vartheta_{21}$ and $\vartheta_{11} < \vartheta_{22}$ if $\sigma^2_{\omega_1} > \sigma^2_{\omega_2}$. ∎

A.11 Instrumental-Variables Regressions with Correlated Measurement Error

This appendix examines the instrumental-variables regression of an individual's test score on both her own and her sibling's schooling when the schooling levels of both siblings are measured with error but both self- and sibling-reported schooling measures are available. The objective here is to characterize the parameter identified by a two-stage least squares regression in which one sibling's reports of the two siblings' schooling are used as instruments for the other sibling's reports of the two siblings' schooling, despite the fact that the measurement errors in the reports are correlated between siblings. The underlying setup is essentially the same as in section 1.2.1 with true schooling and test scores given by equations (1.3) and (1.5), respectively. However, reported schooling is now allowed to differ from true schooling. In particular, let h^i_j denote the report made by sibling $i \in \{1, 2\}$ about the schooling of sibling $j \in \{1, 2\}$. The reported schooling level h^i_j is defined as:

$$h^i_j = s_j + \upsilon^i_j, \tag{A.242}$$

where υ_j^i is the error in sibling i's report about sibling j's schooling. Letting i index one of the siblings and e index the sibling other than i, the error terms υ_i^i and υ_e^i are assumed to have the form:

$$\upsilon_i^i = o_i + u_i^i \text{ and } \upsilon_e^i = o_i + u_e^i, \tag{A.243}$$

where o_1, o_2 have identical variance $\sigma_o^2 > 0$ and correlation coefficient $\rho_o \in (0, 1]$, and both u_1^1, u_2^2 and u_2^1, u_1^2 have identical variance $\sigma_u^2 > 0$. The terms o_1 and o_2, which are intended to represent a shared tendency of siblings from the same family to misreport their schooling levels, are assumed to be correlated between themselves but uncorrelated with (a_1, a_2), (s_1, s_2), (z_1, z_2), (u_1^1, u_2^2), and (u_2^1, u_1^2). For a given $i \in \{1, 2\}$, the terms u_i^i and u_e^i, which represent sibling i's idiosyncratic tendency to misreport her own and her sibling's schooling, can have a correlation $\rho_u \in [-1, 1]$ with each other. However, u_i^i and u_e^i are required to be uncorrelated with (a_1, a_2), (s_1, s_2), and (z_1, z_2) as well as with u_i^e and u_e^e.

The result below analyzes the two-stage least squares regression of sibling i's test score z_i on her own reports h_1^i, h_2^i of the two siblings' schooling levels s_1, s_2 using sibling e's reports h_1^e, h_2^e as instruments for sibling i's reports h_1^i, h_2^i. Specifically, I show that the parameter on sibling i's report of sibling e's schooling h_e^i being estimated by this regression should be less than the regression coefficient on sibling e's schooling s_e in the conditional expectation of sibling i's test score z_i given the two siblings' schooling levels s_1, s_2.

Proposition A.11.1 *Let i index one of the two siblings and e index the sibling other than i. Let $\hat{\theta}_c$ denote the parameter on h_e^i identified by the two-stage least squares regression of sibling i's test score z_i on her own reports h_1^i, h_2^i of both siblings' schooling where sibling e's reports h_1^e, h_2^e are used as instruments for sibling i's reports h_1^i, h_2^i. Let θ_c denote the regression coefficient on s_e in the conditional expectation of sibling i's test score z_i given both siblings' true schooling s_1, s_2. Then $\hat{\theta}_c < \theta_c$.*

Proof The conditional expectation of sibling i's test score z_i given her own and her sibling's true schooling s_i, s_e has the form:

$$\mathbb{E}[z_i | s_i, s_e] = \theta_a + \theta_b s_i + \theta_c s_e. \tag{A.244}$$

Hence, sibling i's test score score z_i can be expressed as:

$$z_i = \theta_a + \theta_b s_i + \theta_c s_e + \varepsilon_i = \theta_a + \theta_b h_i^i + \theta_c h_e^i + \varepsilon_i - (\theta_b \upsilon_i^i + \theta_c \upsilon_e^i), \tag{A.245}$$

where the error term ε_i is independent of s_i and s_e by the definition of the conditional expectation of a normal random variable and uncorrelated with υ_j^i for all i, j by the assumptions made above about the form of the measurement error in schooling. The expressions for θ_b and θ_c in equation (1.6) imply that $\theta_b + \theta_c > 0$.

Next, I analyze the predicted values of sibling i's reports h_i^i, h_e^i given sibling e's reports h_i^e, h_e^e. The coefficient on $(h_i^e, h_e^e)'$ in the linear projection of $(h_i^i, h_e^i)'$ on $(h_i^e, h_e^e)'$ is equal to:

$$\begin{pmatrix} \gamma_b & \gamma_c \\ \gamma_c & \gamma_b \end{pmatrix} = \mathbb{C}\left[\begin{pmatrix} h_i^i \\ h_e^i \end{pmatrix}, \begin{pmatrix} h_i^e \\ h_e^e \end{pmatrix}\right] \mathbb{V}\left[\begin{pmatrix} h_i^e \\ h_e^e \end{pmatrix}\right]^{-1}, \tag{A.246}$$

which can be expressed as:

$$\begin{pmatrix} \gamma_b & \gamma_c \\ \gamma_c & \gamma_b \end{pmatrix} = \begin{pmatrix} \sigma_s^2 + \rho_o \sigma_o^2 & \rho_s \sigma_s^2 + \rho_o \sigma_o^2 \\ \rho_s \sigma_s^2 + \rho_o \sigma_o^2 & \sigma_s^2 + \rho_o \sigma_o^2 \end{pmatrix} \begin{pmatrix} \sigma_s^2 + \sigma_o^2 + \sigma_u^2 & \rho_s \sigma_s^2 + \sigma_o^2 + \rho_u \sigma_u^2 \\ \rho_s \sigma_s^2 + \sigma_o^2 + \rho_u \sigma_u^2 & \sigma_s^2 + \sigma_o^2 + \sigma_u^2 \end{pmatrix}^{-1}. \tag{A.247}$$

Defining the constant $Q > 0$ as:

$$Q = (\sigma_s^2 + \sigma_o^2 + \sigma_u^2)^2 - (\rho_s \sigma_s^2 + \sigma_o^2 + \rho_u \sigma_u^2)^2, \tag{A.248}$$

the parameters γ_b and γ_c are:

$$\begin{aligned} \gamma_b &= Q^{-1}[(\sigma_s^2 + \rho_o \sigma_o^2)(\sigma_s^2 + \sigma_o^2 + \sigma_u^2) - (\rho_s \sigma_s^2 + \rho_o \sigma_o^2)(\rho_s \sigma_s^2 + \sigma_o^2 + \rho_u \sigma_u^2)], \\ \gamma_c &= Q^{-1}[(\rho_s \sigma_s^2 + \rho_o \sigma_o^2)(\sigma_s^2 + \sigma_o^2 + \sigma_u^2) - (\sigma_s^2 + \rho_o \sigma_o^2)(\rho_s \sigma_s^2 + \sigma_o^2 + \rho_u \sigma_u^2)]. \end{aligned} \tag{A.249}$$

Clearly, from equation (A.249), the parameters γ_b and γ_c satisfy $\gamma_b > \gamma_c$ and $\gamma_b + \gamma_c > 0$. Thus, the predicted values of h_i^i, h_e^i given h_i^e, h_e^e have the form:

$$h_i^{i*} = \gamma_p + \gamma_b h_i^e + \gamma_c h_e^e \;\text{ and }\; h_e^{i*} = \gamma_q + \gamma_c h_i^e + \gamma_b h_e^e. \tag{A.250}$$

Now, the parameter on $(h_i^i, h_e^i)'$ identified by the two-stage least squares regression of z_i on $(h_i^i, h_e^i)'$ using $(h_i^e, h_e^e)'$ as an instrument for $(h_i^i, h_e^i)'$ is equal to:

$$\begin{aligned}\begin{pmatrix}\hat{\theta}_b & \hat{\theta}_c\end{pmatrix} &= \mathbb{C}\left[\begin{pmatrix}z_i\end{pmatrix}, \begin{pmatrix}h_i^{i*} \\ h_e^{i*}\end{pmatrix}\right]\mathbb{V}\left[\begin{pmatrix}h_i^{i*} \\ h_e^{i*}\end{pmatrix}\right]^{-1} \\ &= \begin{pmatrix}\theta_b & \theta_c\end{pmatrix} - \mathbb{C}\left[\begin{pmatrix}\theta_b \upsilon_i^i + \theta_c \upsilon_e^i\end{pmatrix}, \begin{pmatrix}\gamma_p + \gamma_b h_i^e + \gamma_c h_e^e \\ \gamma_q + \gamma_c h_i^e + \gamma_b h_e^e\end{pmatrix}\right]\mathbb{V}\left[\begin{pmatrix}h_i^{i*} \\ h_e^{i*}\end{pmatrix}\right]^{-1},\end{aligned} \tag{A.251}$$

where the second equality follows from the decomposition of the test score in equation (A.245) as well as the fact that h_i^{i*} and h_e^{i*} in equation (A.250) are the respective linear projections of h_i^i and h_e^i on h_i^e and h_e^e. The covariance matrix from the second term in the second line of equation (A.251) can be expressed as:

$$\mathbb{C}\left[\begin{pmatrix}\theta_b \upsilon_i^i + \theta_c \upsilon_e^i\end{pmatrix}, \begin{pmatrix}\gamma_p + \gamma_b h_i^e + \gamma_c h_e^e \\ \gamma_q + \gamma_c h_i^e + \gamma_b h_e^e\end{pmatrix}\right] = \rho_o \sigma_o^2 (\theta_b + \theta_c)(\gamma_b + \gamma_c)\begin{pmatrix}1 & 1\end{pmatrix}. \tag{A.252}$$

Hence, defining the constant $P > 0$ as:

$$P = \mathbb{V}(h_i^{i*})\mathbb{V}(h_e^{i*}) - [\mathbb{C}(h_i^{i*}, h_e^{i*})]^2, \tag{A.253}$$

the parameter $\hat{\theta}_c$ in equation (A.251) is given by:

$$\begin{aligned}\hat{\theta}_c &= \theta_c - \rho_o \sigma_o^2 (\theta_b + \theta_c)(\gamma_b + \gamma_c) P^{-1} [\mathbb{V}(h_i^{i*}) - \mathbb{C}(h_i^{i*}, h_e^{i*})] \\ &= \theta_c - \rho_o \sigma_o^2 (\theta_b + \theta_c)(\gamma_b + \gamma_c) P^{-1} (\gamma_b - \gamma_c)^2 [\mathbb{V}(h_e^e) - \mathbb{C}(h_i^e, h_e^e)],\end{aligned} \tag{A.254}$$

where the second equality results after some simplification from substituting for h_i^{i*} and h_e^{i*} using equation (A.250). It follows from the preceding equation that $\hat{\theta}_c < \theta_c$ as desired, because $\theta_b + \theta_c > 0$, $\gamma_b + \gamma_c > 0$, $\gamma_b > \gamma_c$, and $\mathbb{V}(h_e^e) = \mathbb{V}(h_i^e) > \mathbb{C}(h_i^e, h_e^e) = \mathbb{C}(h_e^e, h_i^e)$. ∎

A.12 Comparative Statics for Learning Models

This appendix presents some basic comparative statics results for the learning models from section 1.2. In particular, the proposition below describes how additional signals of one's own or a sibling's performance affect the coefficients on test scores in the regression of one's log wage on one's own and a sibling's test scores and schooling. If learning is individual as in section 1.2.3, then the ratio of the coefficient on a sibling's test score to that on one's own test score remains constant in this regression, regardless of the number of performance signals that a person or her sibling possesses. Of course, under the assumption of individual learning, the number of performance signals available about one's sibling has no effect on any of the coefficients in this regression. By contrast, if learning is social as in section 1.2.4, then this ratio typically increases with more signals about a sibling's performance or fewer signals about one's own performance.

Proposition A.12.1 *Let i index one of the two siblings and e index the sibling other than i. Let $\nu_{ii}(t_i, t_e)$ and $\nu_{ie}(t_i, t_e)$ respectively denote the regression coefficients on the test scores of siblings i and e in the conditional expectation of sibling i's log wage given z_1, z_2 and s_1, s_2, where $t_i > 0$ and $t_e > 0$ are the corresponding numbers of performance signals available about siblings i and e.*

1. *If learning is individual, then $\nu_{ie}(\hat{t}_i, \hat{t}_e)\nu_{ii}(\tilde{t}_i, \tilde{t}_e) = \nu_{ie}(\tilde{t}_i, \tilde{t}_e)\nu_{ii}(\hat{t}_i, \hat{t}_e)$ for all $\hat{t}_i$, $\hat{t}_e$ and $\tilde{t}_i$, $\tilde{t}_e$.*
2. *If learning is social, then:*
   (a) $\nu_{ie}(\hat{t}_i, \hat{t}_e)\nu_{ii}(\tilde{t}_i, \tilde{t}_e) < \nu_{ie}(\tilde{t}_i, \tilde{t}_e)\nu_{ii}(\hat{t}_i, \hat{t}_e)$ whenever $\tilde{t}_e > \hat{t}_e$ and $\tilde{t}_i \leq \hat{t}_i$,
   (b) $\nu_{ie}(\hat{t}_i, \hat{t}_e)\nu_{ii}(\tilde{t}_i, \tilde{t}_e) < \nu_{ie}(\tilde{t}_i, \tilde{t}_e)\nu_{ii}(\hat{t}_i, \hat{t}_e)$ whenever $\tilde{t}_i < \hat{t}_i$ and $\tilde{t}_e \geq \hat{t}_e$.

Proof From equation (1.12), the conditional expectation of sibling i's log wage given s_1, s_2 and z_1, z_2 has the following form under individual learning:

$$\mathbb{E}[\log(w_i)|s_1, s_2, z_1, z_2] = \chi_i(t_i)\mathbb{E}(a_i|s_1, s_2, z_1, z_2) + f_i(s_i, t_i), \tag{A.255}$$

where $\chi_i(t_i)$ is analogous to χ_i in equation (1.9), and $f_i(s_i, t_i)$ is given by equation (1.13). It follows from the preceding equation that:

$$\nu_{ie}(\hat{t}_i, \hat{t}_e)\nu_{ii}(\tilde{t}_i, \tilde{t}_e) = [\chi_i(\hat{t}_i)\pi_f][\chi_i(\tilde{t}_i)\pi_o] = [\chi_i(\tilde{t}_i)\pi_f][\chi_i(\hat{t}_i)\pi_o] = \nu_{ie}(\tilde{t}_i, \tilde{t}_e)\nu_{ii}(\hat{t}_i, \hat{t}_e), \quad \text{(A.256)}$$

where π_o and π_f are the same as the regression parameters on one's own and a sibling's test scores in equation (1.7). This proves the first item in the proposition.

Next, if learning is social, then as in equation (1.23), the conditional expectation of sibling i's log wage given s_1, s_2 and z_1, z_2 has the form:

$$\begin{aligned} \mathbb{E}[\log(v_i)|s_i, s_e, z_i, z_e] =& [1 - \xi_i(t_i, t_e)]\zeta_{ri}(t_e)\mathbb{E}(a_e|s_i, s_e, z_i, z_e) \\ &+ \xi_i(t_i, t_e)\mathbb{E}(a_i|s_i, s_e, z_i, z_e) + p_i(s_i, s_e, t_i), \end{aligned} \quad \text{(A.257)}$$

where $p_i(s_i, s_e, t_i)$ is specified in equation (1.24); $\xi_i(t_i, t_e)$ is defined in equations (1.14) and (1.15) as:

$$\xi_i(t_i, t_e) = t_i\sigma_\eta^{-2}\sigma_{qi}^2(t_i, t_e), \sigma_{qi}^2(t_i, t_e) = [\sigma_{ni}^{-2}(t_e) + t_i\sigma_\eta^{-2}]^{-1}, \sigma_{ni}^2(t_e) = \mathbb{V}(a_i|s_i, s_e, r_e); \quad \text{(A.258)}$$

and ζ_{ri} is given as in equation (A.21) by:

$$\zeta_{ri}(t_e) = \sigma_a^2[\rho_a - \gamma(\delta_o\rho_a + \delta_f)]/(\varsigma^2 + t_e^{-1}\sigma_\eta^2). \quad \text{(A.259)}$$

As explained in the proof of proposition 1.2.4, the bracketed term in the numerator of the preceding equation is positive. Note that $\xi_i(t_i, t_e) \in (0, 1)$ in equation (A.258) and that $\zeta_{ri}(t_e) > 0$ in equation (A.259). In addition, $\xi_i(t_i, t_e)$ is increasing in t_i and non-increasing in t_e, whereas $\zeta_{ri}(t_e)$ is increasing in t_e. From equation (A.257), the regression parameters $\nu_{ii}(t_i, t_e)$ and $\nu_{ie}(t_i, t_e)$ can be expressed as:

$$\begin{aligned} \nu_{ii}(t_i, t_e) &= [1 - \xi_i(t_i, t_e)]\zeta_{ri}(t_e)\pi_f + \xi_i(t_i, t_e)\pi_o, \\ \nu_{ie}(t_i, t_e) &= [1 - \xi_i(t_i, t_e)]\zeta_{ri}(t_e)\pi_o + \xi_i(t_i, t_e)\pi_f, \end{aligned} \quad \text{(A.260)}$$

which is analogous to equation (A.23). Hence, similar to equations (A.24) and (A.25) in the proof of proposition 1.2.4, the statement $\nu_{ie}(\hat{t}_i, \hat{t}_e)\nu_{ii}(\tilde{t}_i, \tilde{t}_e) < \nu_{ie}(\tilde{t}_i, \tilde{t}_e)\nu_{ii}(\hat{t}_i, \hat{t}_e)$ is equivalent to:

$$\begin{aligned}&[1 - \xi_i(\hat{t}_i, \hat{t}_e)]\xi_i(\tilde{t}_i, \tilde{t}_e)\zeta_{ri}(\hat{t}_e)\pi_o^2 + \xi_i(\hat{t}_i, \hat{t}_e)[1 - \xi_i(\tilde{t}_i, \tilde{t}_e)]\zeta_{ri}(\tilde{t}_e)\pi_f^2 \\ &< [1 - \xi_i(\tilde{t}_i, \tilde{t}_e)]\xi_i(\hat{t}_i, \hat{t}_e)\zeta_{ri}(\tilde{t}_e)\pi_o^2 + \xi_i(\tilde{t}_i, \tilde{t}_e)[1 - \xi_i(\hat{t}_i, \hat{t}_e)]\zeta_{ri}(\hat{t}_e)\pi_f^2\end{aligned}. \tag{A.261}$$

Recall from proposition 1.2.2 that $\pi_o^2 > \pi_f^2$. Now, if $\tilde{t}_e > \hat{t}_e$ and $\tilde{t}_i \leq \hat{t}_i$, then $\xi_i(\hat{t}_i, \hat{t}_e) \geq \xi_i(\tilde{t}_i, \tilde{t}_e)$ and $\zeta_{ri}(\hat{t}_e) < \zeta_{ri}(\tilde{t}_e)$. Thus, the inequality in expression (A.261) holds whenever $\tilde{t}_e > \hat{t}_e$ and $\tilde{t}_i \leq \hat{t}_i$, which proves the first part in the second item of the proposition. Next, if $\tilde{t}_i < \hat{t}_i$ and $\tilde{t}_e \geq \hat{t}_e$, then $\xi_i(\hat{t}_i, \hat{t}_e) > \xi_i(\tilde{t}_i, \tilde{t}_e)$ and $\zeta_{ri}(\hat{t}_e) \leq \zeta_{ri}(\tilde{t}_e)$. Thus, the inequality in expression (A.261) holds whenever $\tilde{t}_i < \hat{t}_i$ and $\tilde{t}_e \geq \hat{t}_e$, which proves the second part in the second item of the proposition. ∎

A.13 Simple Model of Employee Referrals

This appendix develops a simple model of employee referrals that seeks to resolve two potential issues mentioned in sections 1.2.4 and 1.6.1. First, the social learning model assumes that one's wage is set equal to the conditional expectation of one's productivity given one's own and a sibling's schooling and performance. If a sibling's characteristics are not observable to a person's employer unless both individuals work for the same firm, then this assumption about wage determination might be unrealistic as a broad description of the labor market. Second, the percentages of individuals obtaining a job through a sibling or also working for the same firm as a sibling are on average moderate in size. If siblings must work for the same firm in order to influence each other's wage, then these percentages might be too small to account for the main estimates of sibling effects.

The model in this section addresses these concerns by relaxing the assumption that one's employer observes the characteristics of one's sibling and by generating an equilibrium with social effects on wages even if siblings work at different firms. The timing of events in this model is similar to that in Montgomery (1991). However, the wage setting process here is different. Specifically, the wage offer made by an informationally advantaged employer is assumed to be observable to other potential employers, who can use this offer to update their beliefs when making counteroffers.[22] In brief, an employer's wage offer may act as a signal to other employers of a worker's productivity.[23]

The basic structure of the model is as follows. There are two siblings and two periods. The siblings differ in seniority with the older and the younger sibling being indexed by 1 and 2, respectively. Each sibling i has a schooling level s_i as well as $p \geq 1$ initial log productivity signals $\{r_{iu}\}_{u=1}^{p}$. In period 1, sibling 1 enters the labor market, whereupon each of $M \geq 2$ firms observes s_1 and $\{r_{1u}\}_{u=1}^{p}$. Each of these firms simultaneously makes a wage offer v_j to sibling 1. Sibling 1 accepts the wage offer of some firm I and works for one period at firm I. Subsequently, firm I observes $q \geq 1$ log output realizations $\{r_{1u}\}_{u=p+1}^{p+q}$ for sibling 1. Having worked, sibling 1 refers sibling 2 to firm I and then leaves the labor market. In period 2, sibling 2 enters the labor market, whereupon firm I observes s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$. Firm I makes a wage offer v_I to sibling 2. Next, $N \geq 2$ other firms observe v_I as well as s_2 and $\{r_{2u}\}_{u=1}^{p}$. Each of these firms simultaneously makes a wage offer v_{Oj} to sibling 2, and sibling 2 accepts a wage offer and works for one period. Subsequently, sibling 2's employer observes $q \geq 1$ log output realizations $\{r_{2u}\}_{u=p+1}^{p+q}$ for sibling 2.

The additional assumptions of the model are as follows. The properties of the variables s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p+q}$ are as described in section 1.2.1.[24] Every wage offer is required to be a positive real number, and each sibling accepts the highest wage offer received.[25] If a firm does not hire a worker in a given period, then the firm obtains a profit of zero for that period. If a firm hires sibling i at wage v in a given period, then the firm obtains a profit of $\exp(\Sigma_{u=p+1}^{p+q} r_{iu}) - v$ for that period, where $\exp(\Sigma_{u=p+1}^{p+q} r_{iu})$ represents sibling i's total output on the job.

The solution concept is perfect Bayesian equilibrium. In period 1, every firm selects v_j so as to maximize the expected discounted value of its profits given the strategies of the other play-

[22]See Barron et al. (2006) for an empirical and theoretical analysis of counteroffers in labor markets. Postel-Vinay and Robin (2002) and Pinkston (2009) present models in which employers observe wage offers from other firms and can extend counteroffers in response.

[23]See Waldman (1984) and Golan (2009), respectively, for models in which job assignments and accepted wages serve as signals of a worker's ability.

[24]In particular, see equations (1.3) and (1.4) for the specifications of s_i and r_{iu}, respectively.

[25]In the treatment here, workers are permitted to use mixed strategies when accepting wage offers, although firms are restricted to use pure strategies when making wage offers. The results of the analysis do not change if firms are allowed to randomize over different wage offers.

ers as well as its beliefs about each sibling i's total output $\exp(\Sigma_{u=p+1}^{p+q} r_{iu})$ conditional on s_1 and $\{r_{1u}\}_{u=1}^{p}$. In period 2, firm I chooses v_I so as to maximize the expected value of its profits given the strategies of the other players in addition to its beliefs about sibling 2's total output $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ conditional on s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$. Each remaining employer then chooses v_{Oj} so as to maximize the expected value of its profits given the strategies of the other players in addition to its beliefs about sibling 2's total output $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ conditional on v_I as well as s_2 and $\{r_{2u}\}_{u=1}^{p}$. Based on the strategies of the players, firms' beliefs are derived from Bayes' rule whenever possible.

In order to solve the model above, I focus on the separating equilibria.[26] The result below establishes the existence of a separating equilibrium. In addition, it shows that in any separating equilibrium, the wage accepted by the older sibling is equal to the conditional expectation of her total output given her own schooling and initial log productivity signals, and the wage accepted by the younger sibling is equal to the conditional expectation of her total output given both siblings' schooling and initial log productivity signals as well as the older sibling's log output realizations.

Proposition A.13.1 *There exists a separating perfect Bayesian equilibrium. In any separating equilibrium, the following hold:*

1. *The wage w_1 accepted by sibling 1 is equal to the conditional expectation of* $\exp(\Sigma_{u=p+1}^{p+q} r_{1u})$ *given s_1 and $\{r_{1u}\}_{u=1}^{p}$.*

2. *The wage w_2 accepted by sibling 2 is equal to the conditional expectation of* $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ *given s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$.*

Proof I begin by providing an example of a separating equilibrium. In period 2, after all the wage offers have been made, sibling 2 accepts the wage offer v_I of firm I if v_I is greater than the highest wage offer $\max_j v_{Oj}$ of the other firms. If v_I is less than or equal to $\max_j v_{Oj}$, then sibling 2 accepts the wage offer v_{Ok} of some firm k other than I that makes an offer of $\max_j v_{Oj}$. If multiple offers by firms other than I are equal to $\max_j v_{Oj}$, then sibling 2 randomly selects an offer, assigning equal probability to each such offer.

[26] To be clear, a separating equilibrium here is a perfect Bayesian equilibrium in which firm I makes a different wage offer v_I to sibling 2 for each of its possible equilibrium beliefs about sibling 2's total output $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ conditional on s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$.

After observing firm I's wage offer v_I to sibling 2, every other firm believes that $\Sigma_{u=p+1}^{p+q} r_{2u}$ is normally distributed with mean $\log(v_I) - \frac{1}{2}\sigma_I^2$ and variance $\sigma_I^2 = \mathbb{V}(\Sigma_{u=p+1}^{p+q} r_{2u} | s_1, s_2, \{r_{1u}\}_{u=1}^{p+q}, \{r_{2u}\}_{u=1}^{p})$. Each of these firms offers sibling 2 a wage v_{Oj} equal to v_I. After observing sibling 1's log output realizations $\{r_{1u}\}_{u=p+1}^{p+q}$, firm I believes that $\Sigma_{u=p+1}^{p+q} r_{2u}$ is normally distributed with mean $\mu_I = \mathbb{E}(\Sigma_{u=p+1}^{p+q} r_{2u} | s_1, s_2, \{r_{1u}\}_{u=1}^{p+q}, \{r_{2u}\}_{u=1}^{p})$ and variance σ_I^2. Firm I offers sibling 2 a log wage $\log(v_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$.

In period 1, after observing sibling 1's schooling s_1 and initial log productivity signals $\{r_{1u}\}_{u=1}^{p}$, every firm believes that $\Sigma_{u=p+1}^{p+q} r_{iu}$ is normally distributed with mean $\mu_{Oi} = \mathbb{E}(\Sigma_{u=p+1}^{p+q} r_{iu} | s_1, \{r_{1u}\}_{u=1}^{p})$ and variance $\sigma_{Oi}^2 = \mathbb{V}(\Sigma_{u=p+1}^{p+q} r_{iu} | s_1, \{r_{1u}\}_{u=1}^{p})$. Each firm offers sibling 1 a log wage $\log(v_j)$ equal to $\mu_{O1} + \frac{1}{2}\sigma_{O1}^2$. Sibling 1 accepts the highest wage offer received $\max_j v_j$. If multiple offers are equal to $\max_j v_j$, then sibling 1 randomly selects an offer, assigning equal probability to each offer.

To see that the strategies and beliefs described above form a separating equilibrium, note first that firm I offers a different log wage $\log(v_I)$ to sibling 2 for each of its possible equilibrium beliefs about $\Sigma_{u=p+1}^{p+q} r_{2u}$ conditional on s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$, where any normal distribution with variance σ_I^2 can be an equilibrium belief. Observe next that the specified beliefs are derived from Bayes' rule. In particular, firm I offers sibling 2 a wage $v_I = \exp(\mu_I + \frac{1}{2}\sigma_I^2)$ equal to the conditional expectation of $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ given s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$. Consequently, upon observing v_I, the other firms use Bayes' rule to infer that $\Sigma_{u=p+1}^{p+q} r_{2u}$ is normally distributed with mean μ_I and variance σ_I^2.

It is now straightforward to confirm that the prescribed strategies are sequentially rational given beliefs. Each sibling always accepts the highest wage offer received. In period 2, every firm obtains an equilibrium expected profit of zero. If a firm other than I were to make an offer greater than its equilibrium offer, then it would obtain a negative expected profit. If such a firm were to make an offer less than its equilibrium offer, then it would obtain an expected profit of zero. If firm I were to make an offer different from its equilibrium offer, then it would continue to receive an expected profit of zero, because the other firms would match this offer, and sibling 2 would never choose to work for firm I. In period 1, each firm obtains an equilibrium expected discounted payoff of zero. If a firm were to offer a lower wage, then it would obtain an expected discounted payoff of zero. If a firm were to offer a higher wage, then it would obtain a negative expected discounted payoff, because it would receive a negative expected profit in period 1 and an expected profit of zero in period 2.

I next show that in any separating equilibrium, the accepted wages must be as given in the statement of the proposition. Suppose that a separating equilibrium is being played. First, if firm I offers sibling 2 a log wage $\log(v_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$, then no other firm k will offer sibling 2 a wage v_{Ok} greater than or equal to v_I unless sibling 2 accepts firm k's offer with probability zero. Thus, if firm I offers sibling 2 a log wage $\log(v_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$, then sibling 2 will accept the offer made by firm I, and firm I will receive a negative expected profit in period 2. However, firm I could obtain an expected payoff of zero in period 2 by instead offering sibling 2 a log wage $\log(v_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$. Hence, there cannot be a separating equilibrium in which firm I offers sibling 2 a log wage $\log(v_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$.

Second, if firm I offers sibling 2 a log wage $\log(v_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$, then no other firm k will make an offer greater than v_I unless sibling 2 accepts firm k's offer with probability zero. Because sibling 2 always accepts the highest wage offer, it must be in such an equilibrium that no firm offers sibling 2 a log wage greater than $\mu_I + \frac{1}{2}\sigma_I^2$ and that sibling 2 receives a log wage of $\mu_I + \frac{1}{2}\sigma_I^2$. Third, if firm I offers sibling 2 a log wage $\log(v_I)$ less than $\mu_I + \frac{1}{2}\sigma_I^2$, then there cannot be an equilibrium in which some firm offers sibling 2 a log wage greater than $\mu_I + \frac{1}{2}\sigma_I^2$. Moreover, some firm must offer sibling 2 a log wage equal to $\mu_I + \frac{1}{2}\sigma_I^2$. Otherwise, there would exist a wage offer $\hat{v}$ greater than $\max(\max_j v_{Oj}, v_I)$ but less than $\exp(\mu_I + \frac{1}{2}\sigma_I^2)$ such that some firm k other than I would have an incentive to deviate by offering sibling 2 the wage $\hat{v}$ instead of making its original wage offer v_{Ok}. Because sibling 2 always accepts the highest wage offer, it must be in such an equilibrium that sibling 2 receives a log wage of $\mu_I + \frac{1}{2}\sigma_I^2$.

Hence, sibling 2's wage must be as specified in the statement of the proposition. Because every firm obtains an expected profit of zero in period 2, the game played in period 1 is equivalent to Bertrand competition among $M \geq 2$ firms making wage offers to sibling 1, where the total expected output from hiring sibling 1 is equal to the conditional expectation of $\exp(\Sigma_{u=p+1}^{p+q} r_{1u})$ given s_1 and $\{r_{1u}\}_{u=1}^{p}$. Consequently, the highest wage offer made to sibling 1 in such an equilibrium is $\exp(\mu_{O1} + \frac{1}{2}\sigma_{O1}^2)$. Hence, sibling 1's wage must be as specified in the statement of the proposition. ∎

Two points should be noted in regard to the result above. First, although attention is restricted to the separating equilibria of the model, other equilibria with different implications for wage setting exist. For example, a pooling equilibrium can be constructed in which the wage accepted by each sibling i is equal to the conditional expectation of her total output $\exp(\Sigma_{u=p+1}^{p+q} r_{iu})$ given her own schooling s_i and initial log productivity signals $\{r_{iu}\}_{u=1}^{p}$. To see this, consider the separating equilibrium described in the first three paragraphs from the proof of proposition A.13.1. Modify the strategies such that firm I always offers sibling 2 a log wage $\log(\hat{v}_I)$ equal to $\hat{\mu}_I + \frac{1}{2}\hat{\sigma}_I^2$, where $\hat{\mu}_I = \mathbb{E}(\Sigma_{u=p+1}^{p+q} r_{2u} | s_2, \{r_{2u}\}_{u=1}^{p})$ and $\hat{\sigma}_I^2 = \mathbb{V}(\Sigma_{u=p+1}^{p+q} r_{2u} | s_2, \{r_{2u}\}_{u=1}^{p})$. Modify the beliefs such that firms other than I believe that $\Sigma_{u=p+1}^{p+q} r_{2u}$ is normally distributed with mean $\hat{\mu}_I$ and variance $\hat{\sigma}_I^2$ whenever firm I offers sibling 2 a log wage of $\hat{\mu}_I + \frac{1}{2}\hat{\sigma}_I^2$. Letting strategies and beliefs be as previously described except for these two changes, it is straightforward to confirm that the resulting strategies and beliefs constitute a pooling equilibrium.[27] In such an equilibrium, each sibling's wage depends only on one's own characteristics.

Second, in the separating equilibrium described in the first three paragraphs from the proof of proposition A.13.1, the older sibling's characteristics always affect the younger sibling's log wage, even though the two siblings never work for the same firm. The reason for this outcome is that the wage offer of the older sibling's former employer reveals private information to other firms about the younger sibling's productivity. Nonetheless, there can also exist a separating equilibrium in which the two siblings always work for the same firm. For example, consider the separating equilibrium described in the first three paragraphs from the proof of proposition A.13.1. Suppose that sibling 2 observes s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$. Modify sibling 2's strategy such that sibling 2 always accepts firm I's offer if firm I makes a wage offer v_I of $\exp(\mu_I + \frac{1}{2}\sigma_I^2)$ and no other firm makes a higher wage offer, where μ_I and σ_I^2 are as defined in the proof of proposition A.13.1. Letting strategies and beliefs be as previously described except for this change, it is straightforward to confirm that the resulting strategies and beliefs constitute a separating equilibrium in which the two siblings always work for the same firm.

A possible concern with the analysis thus far is that the model does not account for the presence of specific human capital. In particular, if sibling 2 acquires human capital specific to firm I while sibling 1 works at firm I, then it would be efficient for sibling 2 to work at firm I as well, because sibling 2's total expected output would be greater at firm I than at any other firm. Therefore, it is unclear if the equilibrium studied earlier would exist when specific human capital is added to the

[27] In addition, various semi-separating equilibria can be constructed.

setup. To address this issue, consider a model identical to the one above except that the total output is $K \cdot \exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ with $K > 1$ if sibling 2 works at firm I. The result below confirms that this model has an equilibrium with the same accepted wages as before. The proof of this claim is simply the first five paragraphs from the proof of proposition A.13.1.

Proposition A.13.2 *In the model with specific human capital, there exists a perfect Bayesian equilibrium such that the following hold:*

1. *The wage w_1 accepted by sibling 1 is equal to the conditional expectation of* $\exp(\Sigma_{u=p+1}^{p+q} r_{1u})$ *given s_1 and $\{r_{1u}\}_{u=1}^{p}$.*

2. *The wage w_2 accepted by sibling 2 is equal to the conditional expectation of* $\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ *given s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$.*

Because the strategies and beliefs described in the first three paragraphs from the proof of proposition A.13.1 are also an equilibrium of the model with specific human capital, it is possible even in the presence of specific human capital for the two siblings to work for different firms. Nonetheless, there can also exist an equilibrium in which the two siblings work for the same firm. For example, consider the separating equilibrium described in the first three paragraphs from the proof of proposition A.13.1. Suppose that sibling 2 observes s_1, s_2 and $\{r_{1u}\}_{u=1}^{p+q}$, $\{r_{2u}\}_{u=1}^{p}$. Modify sibling 2's strategy such that sibling 2 always accepts firm I's offer if firm I makes a wage offer v_I of $\exp(\mu_I + \frac{1}{2}\sigma_I^2)$ and no other firm makes a higher wage offer, where μ_I and σ_I^2 are as defined in the proof of proposition A.13.1. Modify each firm's strategy in period 1 such that each firm's wage offer to sibling 1 is equal to the conditional expectation of $\exp(\Sigma_{u=p+1}^{p+q} r_{1u}) + \delta(K-1)\exp(\Sigma_{u=p+1}^{p+q} r_{2u})$ given s_1 and $\{r_{1u}\}_{u=1}^{p}$, where δ denotes the discount factor between periods. Letting strategies and beliefs be as previously described except for these two changes, it is straightforward to confirm that the resulting strategies and beliefs constitute an equilibrium in which the two siblings always work for the same firm. In addition, the wage accepted by the older sibling in this equilibrium differs from that given in proposition A.13.2. There can also exist an equilibrium in which the wage accepted by the younger sibling differs from that given in proposition A.13.2

The result below shows that if a separating equilibrium is played as in proposition A.13.1, then the wages of the two siblings have the same basic structure as under the social learning model in proposition 1.2.4. That is, if each sibling's log wage is regressed on both siblings' schooling and test scores, then the ratio of the coefficient on a younger sibling's test score to that on one's own test score in an older sibling's log wage is typically lower than the ratio of the coefficient on an older sibling's test score to that on one's own test score in a younger sibling's log wage. Note that the properties of each sibling's test score z_i are as described in section 1.2.1.[28]

Proposition A.13.3 *Suppose that a separating equilibrium is played as in proposition A.13.1. Let ν_{ij} denote the regression coefficient on sibling j's test score in the conditional expectation of sibling i's log wage given s_1, s_2 and z_1, z_2. Then $\nu_{12}\nu_{22} < \nu_{21}\nu_{11}$.*

Proof Using equation (1.12) from the analysis of individual learning in section 1.2.3, the conditional expectation of sibling 1's log wage $\log(w_1)$ given s_1, s_2 and z_1, z_2 has the form:

$$\mathbb{E}[\log(w_1)|s_1, s_2, z_1, z_2] = q\chi_1\mathbb{E}(a_1|s_1, s_2, z_1, z_2) + H_1(s_1), \tag{A.262}$$

where H_1 is some function of s_1, and the parameter χ_1 is defined by:

$$\chi_1 = p\sigma_\eta^{-2}\sigma_{g1}^2, \quad \sigma_{g1}^2 = (\sigma_m^{-2} + p\sigma_\eta^{-2})^{-1}, \quad \sigma_m^2 = \mathbb{V}(a_1|s_1). \tag{A.263}$$

Hence, the coefficients ν_{11} and ν_{12} in the statement of the proposition can be expressed as:

$$\nu_{11} = q\chi_1\pi_o \text{ and } \nu_{12} = q\chi_1\pi_f, \tag{A.264}$$

where π_o and π_f are as defined in equation (1.7). Using equation (1.23) from the analysis of social learning in section 1.2.4, the conditional expectation of sibling 2's log wage $\log(w_2)$ given s_1, s_2 and z_1, z_2 has the form:

$$\begin{aligned}&\mathbb{E}[\log(w_2)|s_1, s_2, z_1, z_2] \\ &= q(1-\xi_2)\zeta_{r2}\mathbb{E}(a_1|s_1, s_2, z_1, z_2) + q\xi_2\mathbb{E}(a_2|s_1, s_2, z_1, z_2) + H_2(s_1, s_2),\end{aligned} \tag{A.265}$$

where H_2 is some function of s_1 and s_2; ζ_{r2} is equal to $(p+q)$ times the coefficient on r_{1u} in the

[28]In particular, see equation (1.5) for the specification of z_i.

conditional expectation of a_2 given s_1, s_2, and $\{r_{1u}\}_{u=1}^{p+q}$; and the parameter ξ_2 is defined by:

$$\xi_2 = p\sigma_\eta^{-2}\sigma_{q2}^2, \quad \sigma_{q2}^2 = (\sigma_{n2}^{-2} + p\sigma_\eta^{-2})^{-1}, \quad \sigma_{n2}^2 = \mathbb{V}(a_2|s_1, s_2, \{r_{1u}\}_{u=1}^{p+q}). \tag{A.266}$$

Hence, the coefficients ν_{21} and ν_{22} in the statement of the proposition can be expressed as:

$$\nu_{21} = q(1-\xi_2)\zeta_{r2}\pi_o + q\xi_2\pi_f \text{ and } \nu_{22} = q(1-\xi_2)\zeta_{r2}\pi_f + q\xi_2\pi_o. \tag{A.267}$$

Note that ζ_{r2} was shown to be positive in the proof of proposition 1.2.4. Now, the statement $\nu_{12}\nu_{22} < \nu_{21}\nu_{11}$ is equivalent to:

$$(q\chi_1\pi_f) \cdot [q(1-\xi_2)\zeta_{r2}\pi_f + q\xi_2\pi_o] < [q(1-\xi_2)\zeta_{r2}\pi_o + q\xi_2\pi_f] \cdot (q\chi_1\pi_o), \tag{A.268}$$

which reduces to $\pi_f^2 < \pi_o^2$. From proposition 1.2.2, we have $\pi_o^2 > \pi_f^2$, completing the proof. ∎

A.14 Within-Family Estimates of AFQT Impacts

Table A.1 reports family fixed-effects estimates of the impact of the AFQT score on the log wage. These results can be interpreted as a simple test for the presence of underlying differences in ability between siblings.[29] For example, if there is a perfect correlation in productive ability a_i among siblings, then any within-family differences in the test score z_i after controlling for schooling s_i should represent test-taking error ω_i, given that the test score has the form in equation (1.5). In this case, assuming that the test-taking error is independent of any variables observable to employers, within-family differences in test scores should be unrelated to within-family variation in log wages, although cross-family differences in test scores, which in part reflect heterogeneity in productive ability, would be associated with cross-family variation in log wages. In a family fixed-effects log wage regression that controls for schooling, the estimated coefficient on the AFQT score is 0.1237 with a standard error of 0.0106. This result suggests that within-family variation in AFQT scores is at least partially attributable to differences in labor market ability between siblings and is not merely an artifact of test-taking error.

[29] In addition, Table A.1 tests for birth order effects on the log wage. As in Kessler (1991), I find no significant evidence of such effects after controlling adequately for age.

Nonetheless, the family fixed-effects estimates of the schooling coefficient in the third and fourth columns of Table A.1 may be biased downward because of measurement error in the schooling variable, which can also lead to inconsistent estimates of the coefficients on the other regressors including the AFQT score. In order to account for measurement error in schooling, one can apply the instrumental-variables procedure in Ashenfelter and Krueger (1994) and Bronars and Oettinger (2006) to the sibling-reported schooling variables in the 1993 sibling roster of the NLSY79. In particular, I difference the regression variables in the third and fourth columns of Table A.1 between the two members of each sibling pair in the main estimation sample during 1993 and use one sibling's report of the difference in schooling levels between the two members as an instrument for the other sibling's report of the schooling difference. The respective two-stage least squares point estimates (standard errors) for the coefficients on schooling are 0.0798 (0.0107) and 0.0565 (0.0122) before and after controlling for the difference in AFQT scores. In the latter specification, the point estimate (standard error) for the coefficient on the AFQT score is 0.1200 (0.0254). The standard errors reported here are adjusted for the correlation in error terms among observations on the same family.

A.15 Further Job Search Estimates

This appendix presents additional empirical results documenting the job search patterns of siblings. Table A.2 displays the percentage of respondents in the NLSY79 who report obtaining their most recent job with the help of a given relative. Of the 52.35 percent of individuals who received assistance from a personal contact, 38.30 percent report that the personal contact is a relative. This relative is a parent in 43.64 percent of cases and a sibling in 20.95 percent of cases. More distant relatives like uncles, aunts, and cousins appear to play a smaller role in helping a person obtain a job than immediate family members. In addition, the percentage of individuals who received help from a sibling varies substantially with sibship size, ranging from 1.30 percent in sibships of size two to 7.20 percent in sibships with at least seven members. Individuals from large families composed of seven or more siblings are more likely to report assistance from a sibling than from a parent. Furthermore, in the majority of instances where a personal contact helped a respondent obtain a job, the contact was initially working for the employer that made a job offer to the respondent. Specifically, in 70.82 percent of cases involving a relative and in 79.52

Table A.1: Family Fixed-Effects Estimates of Impact of AFQT and Schooling on Log Wage

Own AFQT	——	0.1690 (0.0104)	——	0.1237 (0.0106)
Own Schooling	——	——	0.0573 (0.0035)	0.0413 (0.0036)
Birth Order 2	-0.0082 (0.0171)	-0.0065 (0.0161)	-0.0084 (0.0164)	-0.0072 (0.0159)
Birth Order 3	-0.0324 (0.0242)	-0.0195 (0.0226)	-0.0203 (0.0229)	-0.0142 (0.0221)
Birth Order 4	-0.0469 (0.0303)	-0.0296 (0.0286)	-0.0205 (0.0287)	-0.0152 (0.0279)
Birth Order 5	-0.0183 (0.0400)	-0.0107 (0.0376)	-0.0013 (0.0373)	-0.0004 (0.0363)
Birth Order 6	0.0034 (0.0485)	0.0247 (0.0446)	0.0372 (0.0447)	0.0433 (0.0431)
Birth Order 7+	0.0391 (0.0585)	0.0366 (0.0533)	0.0544 (0.0535)	0.0483 (0.0513)
Female	-0.1739 (0.0128)	-0.1797 (0.0120)	-0.2033 (0.0123)	-0.1994 (0.0119)
R^2	0.5308	0.5492	0.5488	0.5572
Families	1993	1993	1993	1993
Individuals	4726	4726	4726	4726
Observations	56552	56552	56552	56552

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for a quartic polynomial in age and a quartic time trend. The sample used here includes any individual who appears in the main estimation sample in some survey year. Each individual is included in every year in which she has left school for the first time, has a wage observation on a full-time job, and has non-missing data on the regressors.

percent of cases involving a sibling, the personal contact worked for the individual's most recent employer.

As a robustness check on the cross-sectional estimates presented in the paper, Table A.3 reports the corresponding family fixed-effects estimates of the influence of birth order on job search behavior. The upper panel estimates the impact of birth order on an individual's likelihood of obtaining her most recent job with the help of a sibling. The point estimate (standard error) for the coefficient on birth order is 0.0172 (0.0047) when all individuals in the sample are included and 0.0145 (0.0060) when individuals still in school are excluded. The lower panel estimates the impact of birth order on an individual's likelihood of obtaining her most recent job with the help of a sibling who was working for the employer that offered her the job. The point estimate (standard error) for the coefficient on birth order is 0.0150 (0.0044) for the entire sample and 0.0114 (0.0056) for those out of school. Estimating a separate coefficient for each birth order up to seven and higher, the use of a sibling in job search appears to increase monotonically with birth order, especially in the full sample. Restricting the sample to individuals who have left school for the first time, later-born children are on average more likely than earlier-born children to obtain a job through a sibling, although there is little indication of such an effect at the upper and lower tails of the birth order distribution.

A.16 AFQT Impacts for Youngest and Oldest Siblings

An alternative to performing the empirical analysis on all the sibling pairs in the labor market would be to focus on the impacts of the youngest and the oldest such siblings from each family, because these individuals have the largest difference in age between them and so may be most likely to have disparate effects on the log wage. Therefore, I construct two datasets, the first of which contains in a given year any person included along with both an older and a younger sibling in the main estimation sample in that year, and the second of which comprises in each year all the sibling pairs in the main estimation sample composed of the oldest and the youngest sibling from a family in the labor market in that year. In Table A.4, the upper panel reports results from regressing a middle sibling's log wage on her own AFQT score as well as those of her youngest and oldest siblings in the labor market, and the lower panel displays estimates from regressions of an oldest sibling's log wage on her own and her youngest sibling's AFQT scores and of a youngest sibling's log wage on her own and her oldest sibling's AFQT scores.

Table A.2: Probability of Given Relative Helping Respondent Obtain Most Recent Job

	Entire Sample	Sibship Size 1	2	3	4	5	6	7+
Percentage Receiving Help from:								
Personal Contact	52.35	49.66	51.67	51.70	53.10	51.77	53.66	53.00
Relative	20.05	12.93	17.66	18.64	20.40	21.20	23.41	21.66
Father	5.28	4.08	7.49	5.77	5.71	4.92	4.39	3.68
Mother	3.47	4.08	3.50	4.33	4.10	3.07	3.66	2.03
Brother	2.08	0.00	0.65	1.29	1.72	2.84	3.17	3.57
Sister	2.12	0.00	0.65	1.34	1.94	2.69	3.05	3.63
Uncle	1.15	1.02	1.14	1.03	1.50	0.84	1.46	1.04
Aunt	0.92	0.34	1.30	0.88	1.00	0.84	1.46	0.55
Cousin	1.39	1.36	1.30	0.82	1.39	1.46	1.46	1.98
Percentage Receiving Help from and Working for Same Employer as:								
Personal Contact	35.06	29.93	35.64	33.83	35.03	35.33	36.71	35.90
Relative	14.20	8.16	12.53	12.20	14.19	15.67	16.83	16.22
Father	3.67	2.72	4.88	3.96	3.88	3.53	3.54	2.64
Mother	2.06	2.38	2.20	2.32	2.49	1.69	2.68	1.21
Brother	1.72	0.00	0.57	1.08	1.55	2.38	1.83	3.08
Sister	1.62	0.00	0.65	0.57	1.61	2.07	2.32	3.02
Uncle	0.85	1.02	0.98	0.72	1.00	0.54	1.34	0.71
Aunt	0.67	0.00	0.90	0.72	0.50	0.77	1.10	0.49
Cousin	1.12	1.02	1.06	0.57	1.11	1.31	1.10	1.65

Note: The tabulations include all 9210 individuals in the NLSY79 with non-missing responses to the relevant questions on job search methods in the 1982 survey. Respondents were first asked, “Was there anyone specifically who helped you get your job with [employer name]?” If so, this question was followed by, “Was this person working for [employer name] when you were first offered this job?” Those answering the first question affirmatively were also asked whether this person was a relative and, if so, what was this person’s relationship to them.

Table A.3: Family Fixed-Effects Estimates of Relationship of Birth Order to Probability of Sibling Helping Respondent Obtain Most Recent Job

	Receive Help from Sibling			
	Entire Sample		Out of School	
Birth Order	0.0172 (0.0047)	——	0.0145 (0.0060)	——
Birth Order 2	——	0.0113 (0.0092)	——	-0.0050 (0.0136)
Birth Order 3	——	0.0457 (0.0150)	——	0.0394 (0.0195)
Birth Order 4	——	0.0778 (0.0192)	——	0.0528 (0.0235)
Birth Order 5	——	0.0807 (0.0266)	——	0.0749 (0.0371)
Birth Order 6	——	0.0902 (0.0331)	——	0.0996 (0.0430)
Birth Order 7+	——	0.1207 (0.0458)	——	0.0812 (0.0552)
R^2	0.4299	0.4318	0.4503	0.4528
Families	1621	1621	946	946
Individuals	3716	3716	2090	2090
	Receive Help from and Work for Same Employer as Sibling			
	Entire Sample		Out of School	
Birth Order	0.0150 (0.0044)	——	0.0114 (0.0056)	——
Birth Order 2	——	0.0091 (0.0088)	——	-0.0056 (0.0134)
Birth Order 3	——	0.0466 (0.0140)	——	0.0425 (0.0184)
Birth Order 4	——	0.0755 (0.0179)	——	0.0486 (0.0218)
Birth Order 5	——	0.0747 (0.0248)	——	0.0591 (0.0346)
Birth Order 6	——	0.0815 (0.0289)	——	0.0844 (0.0396)
Birth Order 7+	——	0.1048 (0.0429)	——	0.0670 (0.0525)
R^2	0.4302	0.4331	0.4440	0.4474
Families	1621	1621	946	946
Individuals	3716	3716	2090	2090

Note: Heteroscedasticity-robust standard errors are reported in parentheses. All specifications include a dummy variable for gender. In the first pair of columns, the sample comprises all sibships in the NLSY79 containing at least two members who have non-missing responses to the job search questions in the 1982 survey. In the second pair of columns, the sample is restricted to those sibships in which at least two members have additionally left school for the first time when surveyed in 1982. Individuals with missing data on birth order are excluded when constructing each estimation sample.

Table A.4: Impact of Own AFQT and AFQT of Youngest and Oldest Sibling in Labor Market on Log Wage

	Middle Siblings in Labor Market					
Oldest Sibling's AFQT	0.0617	0.0440	0.0617	0.0467	0.0526	0.0381
	(0.0241)	(0.0241)	(0.0226)	(0.0214)	(0.0263)	(0.0237)
Youngest Sibling's AFQT	-0.0106	-0.0287	-0.0307	-0.0417	-0.0518	-0.0604
	(0.0266)	(0.0260)	(0.0247)	(0.0232)	(0.0242)	(0.0239)
Own AFQT	0.2229	0.2005	0.1446	0.1262	0.1533	0.1377
	(0.0211)	(0.0233)	(0.0234)	(0.0238)	(0.0233)	(0.0239)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Both Siblings' Schooling	No	No	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT impacts (p-value)	——	——	——	——	0.0105	0.0114
R^2	0.3798	0.4101	0.4139	0.4428	0.4180	0.4462
Families	496	496	496	496	496	496
Individuals	626	626	626	626	626	626
Sibling Trios	828	828	828	828	828	828
Observations	4085	4085	4085	4085	4085	4085
	Oldest and Youngest Siblings in Labor Market					
Oldest Sibling's AFQT $\times$ Youngest Sibling	0.0333	0.0237	0.0206	0.0194	0.0200	0.0193
	(0.0102)	(0.0105)	(0.0101)	(0.0104)	(0.0114)	(0.0116)
Youngest Sibling's AFQT $\times$ Oldest Sibling	0.0111	-0.0032	-0.0018	-0.0082	-0.0178	-0.0210
	(0.0119)	(0.0121)	(0.0116)	(0.0117)	(0.0126)	(0.0126)
Own AFQT $\times$ Youngest Sibling	0.1644	0.1475	0.0990	0.0951	0.0991	0.0951
	(0.0113)	(0.0117)	(0.0118)	(0.0119)	(0.0118)	(0.0120)
Own AFQT $\times$ Oldest Sibling	0.2209	0.2012	0.1579	0.1503	0.1615	0.1543
	(0.0118)	(0.0120)	(0.0127)	(0.0127)	(0.0126)	(0.0126)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0278	0.0231
R^2	0.3241	0.3362	0.3526	0.3595	0.3535	0.3602
Families	1993	1993	1993	1993	1993	1993
Individuals	4684	4684	4684	4684	4684	4684
Sibling Pairs	6064	6064	6064	6064	6064	6064
Observations	37128	37128	37128	37128	37128	37128

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling trio or pair. Included also are indicators for missing data on a given variable, a third-order multivariate polynomial in the ages of the siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the siblings' birth orders. In the lower panel, the coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In each year, the dataset used in the upper panel consists of those individuals appearing in the main estimation sample along with both younger and older siblings. The covariates for the youngest and the oldest of these siblings are used in the analysis. In a given year, the dataset used in the lower panel comprises those sibling pairs in the main estimation sample containing the oldest and the youngest sibling from a family in the labor market.

The estimates in Table A.4 largely confirm the findings in Table 1.3. In the upper panel, an oldest sibling's AFQT score has a greater estimated impact than a youngest sibling's AFQT score on a middle sibling's log wage. Although the estimated coefficient on an oldest sibling's AFQT score is always positive, the coefficient on a youngest sibling's AFQT score often has a negative and significant sign, which can be attributed to a positive correlation in testing error among siblings. Performing the test from proposition A.5.2, the equality of the coefficients on an oldest and a youngest sibling's AFQT scores can be safely rejected at the five percent level of significance, which is inconsistent with individual learning but compatible with social learning. In the lower panel, an oldest sibling's AFQT score has a greater estimated impact on a youngest sibling's log wage than vice versa, and one's own AFQT score has a higher coefficient in the log wage of an oldest sibling than in the log wage of a youngest sibling. Testing the restriction in proposition 1.2.3, the benchmark prediction of the individual learning model can be rejected at the five percent level of significance on the current subsample. In particular, the ratio of the coefficient on an oldest sibling's AFQT score to that on one's own AFQT score in a youngest sibling's log wage is significantly greater than the ratio of the coefficient on a youngest sibling's AFQT score to that on one's own AFQT score in an oldest sibling's log wage. This result accords with the prediction of the social learning model in proposition 1.2.4.

I now provide evidence against the claim that the asymmetric impacts of an oldest and a youngest sibling's AFQT scores on the log wage are due to interactions among siblings prior to labor market entry. I assemble two datasets corresponding to the samples in the upper and lower panels of Table A.4. Using the first of the two datasets, which contains one observation on each sibling trio in the upper panel of Table A.4, I regress a middle sibling's AFQT score and schooling on the AFQT scores of both her youngest and her oldest siblings. The upper panel of Table A.5 displays the results of these regressions. In the first two columns, I find that an oldest sibling's AFQT score has an insignificantly weaker relationship than a youngest sibling's AFQT score to a middle sibling's AFQT score. In addition, the estimates in the third through sixth columns indicate that an oldest sibling's AFQT score has a significantly smaller impact than a youngest sibling's AFQT score on a middle sibling's schooling. Using the second of the two datasets, which contains one observation on each sibling pair in the lower panel of Table A.4, I regress a youngest sibling's AFQT score and schooling on an oldest sibling's AFQT score and vice versa. The lower panel of Table A.5 reports the resulting estimates. A youngest sibling's AFQT score has a greater impact on an oldest sibling's AFQT score and schooling than vice versa. These differences are statistically significant in some specifications. Overall, there is little indication that pre-labor market interactions among siblings are driving the results from the log wage regressions, since any asymmetries in the impacts of siblings' AFQT scores in Table A.5 have the opposite direction from those in Table A.4.[30]

A.17 AFQT Impacts by Number of Siblings

A potential issue with the estimates in Table 1.3 is that the specifications account only for social interactions between the two members of each sibling pair, even though some families contain three or more interviewed siblings. As explained in section 1.2.3, this issue would not lead one to wrongly reject the null hypothesis of individual learning if it is in fact correct. Nonetheless, if employer learning has a social component, then the models being estimated may be slightly misspecified, especially to the extent that some families contain more than two siblings. This appendix, therefore, examines whether the findings change appreciably after accounting more adequately for the presence of additional siblings. I implement the empirical strategy outlined in appendix A.4, which extends the model in sections 1.2.3 and 1.2.4 to include an arbitrary number of siblings in each family. First, I classify each pair of siblings in the main estimation sample in a given year into one of three categories depending on the number of siblings from their family in the labor market in that year.[31] Second, I exclude any observations on the period before all the members of a given sibship have left school for the first time, because the social learning model properly applies to the period after all the siblings from a family have entered the labor market. Third, when regressing a younger sibling's log wage on one's own and an older sibling's AFQT scores and vice versa, I control for the schooling levels of all the members of one's sibship.

Table A.6 presents the estimates obtained from this procedure. The log wage regressions are

[30] Because schooling might exert a causal effect on AFQT scores, appendix A.20 replicates the results in Table A.5 substituting heights for AFQT scores as explanatory variables. Using height as an indicator of cognitive ability, an oldest sibling's height is seen to have an insignificantly smaller impact than a youngest sibling's height on a person's AFQT score and schooling.

[31] Note that these categories are based solely on siblings who are respondents in the NLSY79. Because only those members of a household between 14 and 22 years of age in 1979 were interviewed, siblings who did not meet this age restriction are necessarily omitted from the analysis.

Table A.5: Relationship of Own AFQT and Schooling to AFQT of Youngest and Oldest Sibling

	Middle Siblings					
	AFQT		Schooling			
Oldest Sibling's AFQT	0.3023 (0.0404)	0.2387 (0.0401)	0.4961 (0.1222)	0.2550 (0.1242)	0.0551 (0.1125)	-0.0576 (0.1117)
Youngest Sibling's AFQT	0.3921 (0.0454)	0.3329 (0.0470)	1.0187 (0.1193)	0.7638 (0.1169)	0.4467 (0.1054)	0.3278 (0.1100)
Own AFQT	——	——	——	——	1.4589 (0.1020)	1.3095 (0.1056)
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.2369	0.1917	0.0097	0.0077	0.0260	0.0248
R^2	0.5769	0.6252	0.3507	0.4368	0.5012	0.5441
Families	496	496	496	496	496	496
Individuals	626	626	626	626	626	626
Sibling Trios	828	828	828	828	828	828
	Oldest and Youngest Siblings					
	AFQT		Schooling			
Oldest Sibling's AFQT $\times$ Youngest Sibling	0.5223 (0.0199)	0.4069 (0.0208)	1.0393 (0.0523)	0.6126 (0.0507)	0.3196 (0.0410)	0.1355 (0.0418)
Youngest Sibling's AFQT $\times$ Oldest Sibling	0.5547 (0.0190)	0.4302 (0.0207)	1.1726 (0.0507)	0.7325 (0.0501)	0.4082 (0.0389)	0.2280 (0.0404)
Own AFQT	——	——	——	——	1.3780 (0.0380)	1.1725 (0.0399)
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.0443	0.1333	0.0078	0.0124	0.0391	0.0325
R^2	0.4770	0.5249	0.2318	0.3397	0.4083	0.4558
Families	1993	1993	1993	1993	1993	1993
Individuals	4684	4684	4684	4684	4684	4684
Sibling Pairs	6064	6064	6064	6064	6064	6064

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling trio or pair. Included also are indicators for missing data on a given variable and fixed effects for each of the siblings' years of birth. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the siblings' birth orders. In the lower panel, all specifications include an indicator for whether the respondent is the older or the younger sibling in a given pair. The dataset used in the upper panel contains the first observation on each sibling trio in the main estimation sample consisting of a middle sibling along with her youngest and her oldest sibling in the labor market. The dataset used in the lower panel contains the first observation on every sibling pair in the main estimation sample composed of the oldest and the youngest sibling from a family in the labor market.

performed separately for families with two, three, and four siblings in the labor market. As in Table 1.3, I consistently find that an older sibling's AFQT score has a higher estimated coefficient in a younger sibling's log wage than vice versa and that the coefficient on one's own AFQT score is larger in an older than in a younger sibling's log wage. Moreover, the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is greater than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage, which supports the predictions of the social learning model in appendix A.4. Although the difference between these two ratios is not statistically significant at conventional levels for families with two or four siblings in the labor market, it can be rejected with substantial confidence for families that contain three such siblings.[32] Therefore, I continue to find significant evidence consistent with social learning after modifying the empirical strategy to allow for several siblings in each family.

A.18 AFQT Impacts for Siblings at Same Age Level

An advantage of comparing the log wages of siblings at the same age level instead of at the same point in time is that the coefficient restrictions implied by individual learning can be tested more directly. To perform this comparison, I construct a sample of sibling pairs in which each wage observation on an older sibling at a given age is matched with a wage observation on her younger sibling at that age. In particular, I identify, for a given age level, those individuals in the NLSY79 who, when interviewed at that age, have left school for the first time, have a wage observation on a full-time job, and have a non-twin sibling meeting the preceding two criteria at that age. The sample includes one observation for each age level at which a pair of siblings satisfies the preceding conditions, where an observation in which the older of two siblings appears first is distinct from an observation in which the younger of them appears first.[33] In brief, the dataset used here is analogous to the main estimation sample, except that wage observations on older and

[32]In particular, the two-sided p-values for the hypothesis test are 0.0047 for the third column and 0.0135 for the fourth column.

[33]Because the social learning model properly applies to the period after both siblings in a pair have entered the labor market, I exclude an observation on a pair of siblings at a given age if one sibling has not yet left school for the first time when the other sibling is surveyed at that age.

Table A.6: Impact of Own AFQT and AFQT of Younger or Older Sibling on Log Wage by Number of Siblings in Labor Market

	Two Siblings in Labor Market		Three Siblings in Labor Market		Four Siblings in Labor Market	
Older Sibling's AFQT × Younger Sibling	0.0169	0.0176	0.0293	0.0244	0.0569	0.0417
	(0.0124)	(0.0126)	(0.0161)	(0.0162)	(0.0402)	(0.0254)
Younger Sibling's AFQT × Older Sibling	-0.0126	-0.0171	-0.0393	-0.0380	0.0360	0.0050
	(0.0139)	(0.0141)	(0.0173)	(0.0167)	(0.0289)	(0.0266)
Own AFQT × Younger Sibling	0.1028	0.1005	0.0874	0.0794	0.0774	0.0480
	(0.0125)	(0.0126)	(0.0207)	(0.0210)	(0.0329)	(0.0309)
Own AFQT × Older Sibling	0.1561	0.1502	0.1609	0.1485	0.2030	0.1825
	(0.0131)	(0.0133)	(0.0200)	(0.0199)	(0.0335)	(0.0288)
Own Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Every Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.1097	0.0712	0.0047	0.0135	0.3737	0.1572
Families	1966	1966	521	521	94	94
Individuals	4374	4374	1672	1672	394	394
Sibling Pairs	5088	5088	3588	3588	1248	1248
Observations	28176	28176	17630	17630	5628	5628

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of each individual and all of her siblings in the labor market. Included also are indicators for missing data on a given variable, a third-order multivariate polynomial in the ages of all the members of a sibship in the labor market, and a quartic time trend. Family background controls are indicator variables for mother's education, father's education, mother's age, father's age, and one's own and each sibling's birth order. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In order to calculate the estimates reported here, each pair of siblings in the main estimation sample in a given year is classified into a group depending on the number of siblings from their family in the labor market in that year. The analysis is restricted to observations on the period after all siblings from a family have left school for the first time. There are insufficiently many observations on families with five or more siblings in the labor market to compute estimates for them.

younger siblings are matched at a given age level instead of in a given survey year.

I regress the log wage of the older sibling in a pair on her own AFQT score and that of the younger sibling in the pair and the log wage of the younger sibling in a pair on her own AFQT score and that of the older sibling in the pair. Table A.7 reports the results of these regressions. The pattern of results in Table A.7 is similar to that in Table 1.3. An older sibling's AFQT score has a greater estimated impact on a younger sibling's log wage than vice versa, and the estimated coefficient on one's own AFQT score is larger for an older than for a younger sibling. An older sibling has a lower estimated coefficient than a younger sibling on her own schooling, and the estimated coefficient on a younger sibling's schooling in an older sibling's log wage is higher than vice versa. The observed differences between the estimated coefficients on test scores and schooling levels appear to be inconsistent with a simple model of individual learning, which would typically predict that these coefficients should be the same for younger and older siblings at similar age levels, provided that younger and older siblings have symmetric processes of human capital accumulation, the wage structure stays essentially constant over time, and the speed of employer learning does not differ between the older and the younger members of a sibship.

Using the estimates in the last two columns to test the predictions of the learning models in proposition A.2.1, one can individually reject both the hypothesis that the coefficient on an older sibling's AFQT score in a younger sibling's log wage is equal to the coefficient on a younger sibling's AFQT score in an older sibling's log wage and the hypothesis that the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is equal to the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage.[34] Thus, the coefficient restrictions implied by individual learning fail to hold. Consistent with social learning, the coefficient on an older sibling's AFQT score in a younger sibling's log wage is significantly greater than vice versa. Moreover, the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is significantly greater than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage.

[34]The two-sided p-values for the test of equality between the coefficients on an older and a younger sibling's AFQT scores are 0.0064 in the fifth column and 0.0094 in the sixth column. For the test of the hypothesis that the ratio of the coefficient on a sibling's AFQT score to that on one's own AFQT score is the same for older and for younger siblings, the respective p-values are 0.0090 and 0.0120 in the fifth and sixths columns.

Table A.7: Impact of Own AFQT and AFQT of Younger or Older Sibling in Labor Market on Log Wage at Same Age Level

Older Sibling's AFQT × Younger Sibling	0.0402 (0.0110)	0.0255 (0.0104)	0.0317 (0.0113)	0.0254 (0.0104)	0.0296 (0.0124)	0.0245 (0.0116)
Younger Sibling's AFQT × Older Sibling	0.0194 (0.0104)	0.0070 (0.0104)	0.0081 (0.0103)	0.0022 (0.0101)	-0.0122 (0.0114)	-0.0148 (0.0113)
Own AFQT × Younger Sibling	0.1707 (0.0117)	0.1502 (0.0117)	0.1035 (0.0118)	0.0965 (0.0117)	0.1037 (0.0118)	0.0966 (0.0117)
Own AFQT × Older Sibling	0.2016 (0.0115)	0.1829 (0.0116)	0.1496 (0.0127)	0.1402 (0.0121)	0.1551 (0.0125)	0.1464 (0.0120)
Own Schooling × Younger Sibling	——	——	0.0505 (0.0044)	0.0477 (0.0044)	0.0500 (0.0046)	0.0476 (0.0045)
Own Schooling × Older Sibling	——	——	0.0385 (0.0042)	0.0364 (0.0044)	0.0325 (0.0044)	0.0316 (0.0046)
Older Sibling's Schooling × Younger Sibling	——	——	——	——	0.0016 (0.0042)	0.0007 (0.0042)
Younger Sibling's Schooling × Older Sibling	——	——	——	——	0.0165 (0.0043)	0.0153 (0.0042)
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT impacts (p-value)	——	——-	——	——	0.0064	0.0094
test for equality between own AFQT impacts (p-value)	——	——-	——	——	0.0017	0.0018
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0090	0.0120
R^2	0.3101	0.3244	0.3360	0.3456	0.3376	0.3469
Families	1887	1887	1887	1887	1887	1887
Individuals	4481	4481	4481	4481	4481	4481
Sibling Pairs	6630	6630	6630	6630	6630	6630
Observations	36656	36656	36656	36656	36656	36656

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. For a given age level, the dataset comprises those individuals in the NLSY79 who, when interviewed at that age, have left school for the first time, have non-missing data on their AFQT score and schooling, have a valid wage observation on a full-time job, have non-missing sibling data including birth order and sibship size, and have a non-twin sibling also meeting these criteria. The sample includes one observation for every age level at which a pair of siblings satisfies the preceding conditions. An observation on a pair of siblings at a given age level is excluded from the analysis if one sibling has not yet left school for the first time when the other sibling is interviewed at that age.

Three further deviations from a simple model of individual learning are evident in the fifth and sixth columns. First, the coefficient on one's own AFQT score is significantly greater in an older than in a younger sibling's log wage. Second, the impact of an older sibling's schooling on a younger sibling's log wage is significantly lower than vice versa. Third, the coefficient on one's own schooling is significantly smaller in an older than in a younger sibling's log wage.[35] Although the social learning model does not guarantee that these three outcomes will simultaneously arise for every possible parameter value, there do exist parameter values for which the social learning model generates all of these features of the data.

One final point should be made in this context. When comparing the AFQT scores of younger and older siblings in section 1.6.2, I noted that older siblings may have a higher variance in testing error than younger siblings after conditioning on various background variables. This observation raised the question of whether differences between younger and older siblings in the distribution of testing error could explain the relevant patterns in the wage data, even if employer learning were in fact individual. The analysis in appendix A.10 shows that if the error term in test scores has a higher variance for an older than for a younger sibling, then assuming that learning is individual, the coefficient on one's own test score will be smaller in the log wage of an older compared to a younger sibling, and the impact of an older sibling's test score on a younger sibling's log wage will be lower than vice versa. However, as can be seen from the estimates in the fifth and sixth columns of Table A.7, I find significant effects in the opposite direction from those suggested by this explanation. Therefore, it is unlikely that such differences in the composition of siblings' test scores are driving the important asymmetries in the results from the log wage regressions.

A potential issue with the estimates in Table A.7 is that the wage structure may change over time. In particular, if the rates of return to labor market skills are not constant across survey years, then some of the asymmetries between younger and older siblings in the coefficients on one's own and a sibling's test scores and schooling might be attributable to differences in the returns to skills between the times when an older and a younger sibling reach a given age.[36] In order to address

[35]The respective two-sided p-values in the fifth and sixth columns are 0.0017 and 0.0018 for the test of equality between the coefficients on one's own AFQT score, 0.0129 and 0.0120 for the test of equality between the coefficients on a sibling's schooling, and 0.0060 and 0.0106 for the test of equality between the coefficients on one's own schooling.

[36]For example, Katz and Murphy (1992) and Murphy and Welch (1992) document an increase in educational wage differentials during the 1980's. If the returns to skills tend to be rising over time, then a younger sibling's wage at a given age might place greater weight on human capital measures than an older sibling's wage at that age.

this issue, Table A.8 provides estimates for the six specifications in Table A.7 after additionally controlling for interactions between the quartic time trend and both one's own and a sibling's AFQT scores. In addition, the third through sixth columns include interaction terms between the time trend and one's own schooling, and the fifth and sixth columns also contain interaction terms between the time trend and a sibling's schooling.

The pattern of results in Table A.8 is similar to that in Table A.7. Specifically, the point estimates suggest that the log wage of an older sibling at a given age exhibits a lower coefficient on a sibling's AFQT score, a higher coefficient on one's own AFQT score, a higher coefficient on a sibling's schooling, and a lower coefficient on one's own schooling than the log wage of a younger sibling at the same age. Based on the estimates in the fifth and sixth columns, all of these differences, except for the last finding, are statistically significant at either the one or five percent level.[37] These deviations from the restrictions implied by individual learning can all be generated for some parameter values by a model with social learning. Finally, consistent with the predictions of social learning, I find that the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is significantly greater than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage.[38] Thus, the estimates in Table A.8 suggest that the asymmetries between the log wages of younger and older siblings in Table A.7 are unlikely to be explained by changes in the wage structure over time.

Another potential issue with the estimates in Table A.7 is that the employer learning process might differ between older and younger siblings, making the comparison of log wages at the same age level somewhat imprecise. For example, suppose that older siblings obtain more schooling on average than younger siblings. One possibility is that individuals acquire performance signals at a slower rate while in school than while employed; so that, at a given age, older siblings would tend to have fewer signals of their own performance than younger siblings. Another possibility is that individuals obtain ability signals more rapidly while attending school than after entering

[37]The respective two-sided p-values in the fifth and sixth columns are 0.0170 and 0.0229 for the test of equality between the coefficients on a sibling's AFQT score, 0.0003 and 0.0003 for the test of equality between the coefficients on one's own AFQT score, 0.0390 and 0.0344 for the test of equality between the coefficients on a sibling's schooling, and 0.2057 and 0.3121 for the test of equality between the coefficients on one's own schooling.

[38]The two-sided p-values for the test of equality between the ratios of the coefficient on a sibling's test score to that on one's own test score are 0.0137 and 0.0165 for the fifth and sixth columns, respectively.

Table A.8: Impact of Own AFQT and AFQT of Younger or Older Sibling on Log Wage at Same Age Level after Controlling for Interactions of AFQT and Schooling with Time Trend

Older Sibling's AFQT × Younger Sibling	0.0459 (0.0127)	0.0311 (0.0120)	0.0367 (0.0130)	0.0302 (0.0120)	0.0380 (0.0145)	0.0330 (0.0136)
Younger Sibling's AFQT × Older Sibling	0.0293 (0.0121)	0.0167 (0.0120)	0.0162 (0.0121)	0.0100 (0.0118)	0.0015 (0.0138)	-0.0015 (0.0136)
Own AFQT × Younger Sibling	0.1721 (0.0129)	0.1522 (0.0128)	0.1043 (0.0130)	0.0975 (0.0129)	0.1034 (0.0129)	0.0964 (0.0129)
Own AFQT × Older Sibling	0.2236 (0.0130)	0.2054 (0.0129)	0.1596 (0.0145)	0.1504 (0.0138)	0.1634 (0.0144)	0.1548 (0.0138)
Own Schooling × Younger Sibling	——	——	0.0483 (0.0048)	0.0456 (0.0048)	0.0490 (0.0051)	0.0468 (0.0049)
Own Schooling × Older Sibling	——	——	0.0445 (0.0050)	0.0430 (0.0051)	0.0408 (0.0053)	0.0403 (0.0054)
Older Sibling's Schooling × Younger Sibling	——	——	——	——	-0.0018 (0.0046)	-0.0028 (0.0047)
Younger Sibling's Schooling × Older Sibling	——	——	——	——	0.0106 (0.0051)	0.0095 (0.0050)
Own AFQT × Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Sibling's AFQT × Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Own Schooling × Time Trend	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling × Time Trend	No	No	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT impacts (p-value)	——	——	——	——	0.0170	0.0229
test for equality between own AFQT impacts (p-value)	——	——	——	——	0.0003	0.0003
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0137	0.0165
R^2	0.3203	0.3339	0.3468	0.3563	0.3482	0.3574
Families	1887	1887	1887	1887	1887	1887
Individuals	4481	4481	4481	4481	4481	4481
Sibling Pairs	6630	6630	6630	6630	6630	6630
Observations	36656	36656	36656	36656	36656	36656

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. The base year for the time trend is 1993.

the labor market; so that, at a given age, older siblings would tend to have more signals of their own productivity than younger siblings.[39] As a result, even if employer learning is individual, some asymmetries between older and younger siblings in the composition of log wages might persist when comparing siblings at the same age. However, these two possibilities are not very likely to explain the main patterns observed in Table A.7. In particular, assuming that ability is a unidimensional factor, proposition 1.2.3 indicates that if learning is individual, then the ratio of the coefficient on a sibling's test score to that on one's own test score should be independent of the number of performance signals that an individual possesses. By contrast, Table A.7 reveals significant differences in this ratio between older and younger siblings.

As a further check that differences in the rate of employer learning related to differences in schooling are not primarily responsible for the asymmetries in Table A.7, I estimate the specifications in the fifth and sixth columns for three subsamples of the dataset in which both members of a sibling pair have similar amounts of schooling. These subsamples comprise sibling pairs in which both members have less than twelve, exactly twelve, and more than twelve years of schooling. Table A.9 displays the regression output for each group. The estimated impact of an older sibling's AFQT score on a younger sibling's log wage is greater than vice versa, and the estimated coefficient on one's own AFQT score is greater in the log wage of an older compared to a younger sibling. Furthermore, the observed ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is greater than the observed ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage. Although these differences are not statistically significant for every group, the relevant null hypotheses can usually be rejected at the five percent level of significance for siblings with exactly twelve years of schooling.[40] Overall, there is little evidence to suggest that the potential effect of schooling on employer learning is generating the asymmetries in Table A.7.

[39] In particular, Arcidiacono et al. (2010) argue that college attendance hastens the revelation of productive ability to the labor market.

[40] For siblings with exactly twelve years of schooling, the two-sided p-values in the third and fourth columns are respectively 0.0317 and 0.0453 for the test of equality between the coefficients on a sibling's AFQT score, 0.0390 and 0.0505 for the test of equality between the coefficients on one's own AFQT score, and 0.0344 and 0.0518 for the test of equality between the ratios of the coefficient on a sibling's AFQT score to that on one's own AFQT score.

Table A.9: Impact of Own AFQT and AFQT of Younger or Older Sibling on Log Wage at Same Age Level by Years of Schooling Completed

	Less than Twelve Years		Equal to Twelve Years		More than Twelve Years	
Older Sibling's AFQT × Younger Sibling	0.0367	0.0377	0.0330	0.0302	0.0249	0.0212
	(0.0409)	(0.0394)	(0.0210)	(0.0206)	(0.0205)	(0.0207)
Younger Sibling's AFQT × Older Sibling	-0.0834	-0.0843	-0.0330	-0.0284	-0.0254	-0.0236
	(0.0623)	(0.0672)	(0.0232)	(0.0221)	(0.0217)	(0.0210)
Own AFQT × Younger Sibling	0.0879	0.0412	0.1099	0.1025	0.1137	0.1039
	(0.0627)	(0.0611)	(0.0238)	(0.0235)	(0.0223)	(0.0215)
Own AFQT × Older Sibling	0.1402	0.1558	0.1727	0.1600	0.1532	0.1456
	(0.0545)	(0.0511)	(0.0214)	(0.0213)	(0.0233)	(0.0235)
Own Schooling	Yes	Yes	No	No	Yes	Yes
Sibling's Schooling	Yes	Yes	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality between sibling AFQT impacts (p-value)	0.1289	0.1382	0.0317	0.0453	0.0887	0.1352
test for equality between own AFQT impacts (p-value)	0.5552	0.1813	0.0390	0.0505	0.2099	0.1864
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.1867	0.2737	0.0344	0.0518	0.0993	0.1567
Families	261	261	536	536	617	617
Individuals	577	577	1214	1214	1327	1327
Sibling Pairs	730	730	1574	1574	1746	1746
Observations	3248	3248	8232	8232	8726	8726

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In order to compute the estimates reported here, three groups of sibling pairs were identified depending on whether both siblings have completed less than twelve, exactly twelve, or more than twelve years of schooling.

A.19 AFQT Impacts Controlling for Schooling at Time of Test

This appendix examines whether the causal effect of education on test scores is likely to contribute to the disparate impacts of older and younger siblings' AFQT scores on the log wage. As explained in appendix A.3, it may be possible to largely account for the influence of schooling on test scores in Table 1.3 by simply controlling for one's schooling at the time of AFQT administration. Table A.10 documents the relationship between one's AFQT score and both one's own and a sibling's schooling at AFQT administration and when in the labor market. The first two columns present estimates for the specification analyzed in proposition 1.2.1, in which schooling at test administration is assumed to be the same as schooling when in the labor market. Not surprisingly, the coefficient on one's own schooling is significantly positive, which could be due either to a positive causal effect of schooling on the AFQT score or to the positive partial correlation between one's ability and one's own schooling. Furthermore, the coefficient on a sibling's schooling is positive. In the framework in section 1.2.2, this result suggests that the sibling correlation in ability is greater than that in schooling, especially if schooling itself is measured without error or if the measurement error in schooling is highly correlated among siblings.[41]

The third and fourth columns of Table A.10 add one's own and a sibling's schooling at AFQT administration to the list of regressors.[42] The pattern of results in the third and fourth columns is consistent with the assumptions in appendix A.3. First, the coefficients on a person's schooling at labor market entry and at AFQT administration are both significantly positive. In the framework in appendix A.3, the coefficient on one's own schooling at labor market entry reflects the positive relationship between ability and schooling, and the coefficient on one's own schooling at AFQT administration represents the positive causal effect of schooling on the AFQT score. Second, the coefficient on a sibling's schooling at labor market entry is significantly positive, possibly indicating that ability is more highly correlated among siblings than schooling. Third, the coefficient on a sibling's schooling at AFQT administration is statistically insignificant and close to zero. This finding is consistent with the assumption in appendix A.3 that a sibling's schooling at AFQT ad-

[41] In appendix A.22, I examine the issue of measurement error in more detail, presenting evidence that this finding does not seem to be due simply to error in the schooling reports.

[42] As in Neal and Johnson (1996), a respondent's schooling at the time of taking the AFQT is treated as being the highest grade completed as of May 1980, because the ASVAB was administered to participants in the NLSY79 during the summer and fall of 1980.

Table A.10: Relationship of Own and Sibling's Schooling at Labor Market Entry and at AFQT Administration to AFQT Score

Sibling's Schooling at Labor Market Entry	0.0483	0.0262	0.0468	0.0238
	(0.0050)	(0.0052)	(0.0063)	(0.0065)
Own Schooling at Labor Market Entry	0.1984	0.1733	0.1684	0.1441
	(0.0053)	(0.0057)	(0.0069)	(0.0071)
Sibling's Schooling at Time of AFQT			-0.0009	0.0022
	——	——	(0.0105)	(0.0099)
Own Schooling at Time of AFQT			0.0867	0.0847
	——	——	(0.0133)	(0.0127)
Family Background Controls	No	Yes	No	Yes
R^2	0.5333	0.5598	0.5400	0.5660
Families	1993	1993	1956	1956
Individuals	4726	4726	4631	4631
Sibling Pairs	7074	7074	6904	6904

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable and fixed effects for each of the two siblings' years of birth. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The dataset contains the first observation on every sibling pair appearing in the main estimation sample. For each pair of siblings, schooling at labor market entry is the highest grade completed as of the survey year in which the pair first appears in the main estimation sample, and schooling at time of AFQT is the highest grade completed as of May 1980, being that the AFQT was administered during the summer and fall of 1980.

ministration is unrelated to one's underlying ability and testing error after controlling for one's own and a sibling's schooling when in the labor market as well as one's own schooling at AFQT administration.

Table A.11 presents estimates for the six specifications in Table 1.3 after including both one's own and a sibling's schooling at AFQT administration as regressors. The basic pattern of results is unchanged. The estimated impact of an older sibling's AFQT score on a younger sibling's log wage is greater than vice versa, and the estimated coefficient on one's own AFQT score is higher in the log wage of an older compared to a younger sibling. Each of these differences is statistically significant at least at the five percent level. Using the estimates in the fifth and sixth columns, the restriction imposed by the individual learning model in proposition A.3.1 can be safely rejected.[43] Consistent with the implications of social learning, the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is significantly greater than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage. Hence, the asymmetric impacts of older and younger siblings' AFQT scores in Table 1.3 do not appear to be due to the causal effect of schooling on AFQT scores.

As explained in appendix A.3, an additional test of the restriction that one's own and a sibling's schooling levels at AFQT administration are uninformative about a person's ability after controlling for one's own and a sibling's schooling levels when in the labor market can be constructed by regressing a person's log wage on one's own and a sibling's schooling both at AFQT administration and when in the labor market. In particular, the coefficients on both one's own and a sibling's schooling at the time of the AFQT should be zero in this regression. Table A.12 reports the results of such a test. The dataset used here is obtained by restricting the main estimation sample to the period before either member of each sibling pair acquires more education after initially leaving school. The rationale for excluding observations following a return to school is that if one's educational attainment changes after entering the labor market, then one's schooling at AFQT administration might be acting as a proxy for one's education upon first leaving school, which could be correlated with one's ability despite conditioning on one's current level of schooling.

[43] The two-sided p-values for this test are 0.0043 and 0.0055 for the fifth and sixth columns, respectively.

Table A.11: Impact of Own AFQT and AFQT of Younger or Older Sibling in Labor Market on Log Wage after Controlling for Own and Sibling's Schooling at Time of AFQT Administration

Older Sibling's AFQT × Younger Sibling	0.0318	0.0204	0.0314	0.0256	0.0286	0.0233
	(0.0115)	(0.0110)	(0.0114)	(0.0108)	(0.0120)	(0.0114)
Younger Sibling's AFQT × Older Sibling	0.0012	-0.0096	-0.0031	-0.0096	-0.0180	-0.0218
	(0.0106)	(0.0104)	(0.0105)	(0.0102)	(0.0113)	(0.0112)
Own AFQT × Younger Sibling	0.1538	0.1373	0.0960	0.0902	0.0968	0.0910
	(0.0123)	(0.0120)	(0.0119)	(0.0118)	(0.0118)	(0.0117)
Own AFQT × Older Sibling	0.1967	0.1822	0.1639	0.1541	0.1671	0.1578
	(0.0125)	(0.0122)	(0.0130)	(0.0125)	(0.0129)	(0.0124)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	No	No	Yes	Yes
Own Schooling at Time of AFQT	Yes	Yes	Yes	Yes	Yes	Yes
Sibling's Schooling at Time of AFQT	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0043	0.0055
R^2	0.3369	0.3494	0.3590	0.3680	0.3600	0.3688
Families	1956	1956	1956	1956	1956	1956
Individuals	4631	4361	4361	4361	4361	4361
Sibling Pairs	6904	6904	6904	6904	6904	6904
Observations	53462	53462	53462	53462	53462	53462

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. The main estimation sample described in the text is used here.

The first two columns of Table A.12 report the results of regressing one's log wage on one's own and a sibling's schooling when in the labor market. Similar to the corresponding findings in Table A.10, where the dependent variable is the AFQT score, the coefficients on one's own and a sibling's schooling when in the labor market are both positive and significant at the one percent level. As explained in section 1.2.2, the positive coefficient on a sibling's schooling can possibly be interpreted as indicating that ability is more highly correlated among siblings than schooling. The third and fourth columns of Table A.12 add one's own and a sibling's schooling at AFQT administration to the specifications in the first and second columns. As before, the coefficients on one's own and a sibling's schooling when in the labor market are both significantly positive. Moreover, the coefficient on a sibling's schooling at the time of the AFQT is close to zero in magnitude and far from being statistically significant, which is analogous to the results in Table A.10 when using the AFQT score instead of the log wage as the dependent variable. Finally, unlike in Table A.10, where one's own schooling at the time of the AFQT is more likely to have a causal effect on the dependent variable, the coefficient on one's own schooling at AFQT administration is relatively small in size and not significantly different from zero. Overall, the results in Table A.12 are consistent with the assumptions in appendix A.3 about the relationship of one's ability to one's own and a sibling's schooling at AFQT administration.

For completeness, Table A.13 presents estimates for the six specifications in Table A.11 using the sample from Table A.12. The basic pattern of results in Table A.13 is the same as that in Table A.11. The estimated coefficient on one's own AFQT score is larger for an older than for a younger sibling, and the estimated impact of an older sibling's AFQT score on a younger sibling's log wage is greater than vice versa. Furthermore, in the fifth and sixth columns, the ratio of the coefficient on an older sibling's AFQT score to that on one's own AFQT score in a younger sibling's log wage is significantly higher than the ratio of the coefficient on a younger sibling's AFQT score to that on one's own AFQT score in an older sibling's log wage.[44] Thus, I continue to find evidence contrary to individual learning but consistent with social learning. The causal effect of schooling on AFQT scores does not seem to play a role in generating the relevant asymmetries in Table 1.3.

[44]The two-sided p-values for the test of equality between these ratios is 0.0310 for the fifth and 0.0259 for the sixth column.

Table A.12: Relationship of Own and Sibling's Schooling at Labor Market Entry and at AFQT Administration to Log Wage after Excluding Observations on Sibling Pairs in Which Either Member Has Returned to School

Sibling's Schooling at Labor Market Entry	0.0194	0.0150	0.0185	0.0145
	(0.0034)	(0.0039)	(0.0043)	(0.0045)
Own Schooling at Labor Market Entry	0.0729	0.0659	0.0694	0.0638
	(0.0039)	(0.0043)	(0.0050)	(0.0052)
Sibling's Schooling at Time of AFQT			0.0027	0.0022
	——	——	(0.0065)	(0.0063)
Own Schooling at Time of AFQT			0.0080	0.0048
	——	——	(0.0088)	(0.0084)
Family Background Controls	No	Yes	No	Yes
R^2	0.3066	0.3186	0.3072	0.3193
Families	1454	1454	1430	1430
Individuals	3363	3363	3302	3302
Sibling Pairs	4636	4636	4534	4534
Observations	30276	30276	29686	29686

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In each survey year, the sample used here contains all observations in the main estimation sample, except for those in which either member of a sibling pair has completed more years of schooling when interviewed in that year than when leaving school for the first time.

Table A.13: Impact of Own AFQT and AFQT of Younger or Older Sibling in Labor Market on Log Wage after Controlling for Schooling at AFQT Administration and Excluding Observations Following a Return to School

Older Sibling's AFQT × Younger Sibling	0.0377	0.0312	0.0314	0.0313	0.0270	0.0267
	(0.0144)	(0.0142)	(0.0147)	(0.0142)	(0.0158)	(0.0152)
Younger Sibling's AFQT × Older Sibling	-0.0027	-0.0081	-0.0067	-0.0083	-0.0200	-0.0195
	(0.0126)	(0.0127)	(0.0128)	(0.0128)	(0.0142)	(0.0141)
Own AFQT × Younger Sibling	0.1402	0.1273	0.0899	0.0850	0.0904	0.0859
	(0.0145)	(0.0148)	(0.0147)	(0.0148)	(0.0147)	(0.0147)
Own AFQT × Older Sibling	0.1944	0.1746	0.1721	0.1583	0.1741	0.1610
	(0.0151)	(0.0150)	(0.0163)	(0.0159)	(0.0162)	(0.0158)
Own Schooling	No	No	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	No	No	Yes	Yes
Own Schooling at Time of AFQT	Yes	Yes	Yes	Yes	Yes	Yes
Sibling's Schooling at Time of AFQT	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	——	——	——	——	0.0310	0.0259
R^2	0.3235	0.3375	0.3416	0.3529	0.3429	0.3539
Families	1430	1430	1430	1430	1430	1430
Individuals	3302	3302	3302	3302	3302	3302
Sibling Pairs	4534	4534	4534	4534	4534	4534
Observations	29686	29686	29686	29686	29686	29686

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. In each survey year, the sample used here contains all observations in the main estimation sample, except for those in which either member of a sibling pair has completed more years of schooling when interviewed in that year than when leaving school for the first time.

A.20 Sibling Height Impacts

A potential issue with the specifications in Tables 1.5 and A.5 is that the AFQT scores used as explanatory variables might be endogenously influenced by educational investments. To address this concern, I reproduce the earlier analysis replacing the AFQT score with height as a measure of cognitive skills.[45] As noted in Case and Paxson (2008), a person's adult height is partly determined by nutritional conditions in early childhood and is positively associated with intellectual ability. Because schooling is unlikely to have a large causal effect on a person's height, the use of heights instead of AFQT scores reduces the likelihood that a person's education would alter the ability measures used as regressors.

Table A.14, which is an analogue of Table 1.5, reports the results of regressing a younger (resp. older) sibling's AFQT score and schooling on her own and her older (resp. younger) sibling's heights. In all cases, the coefficient on an older sibling's height is insignificantly smaller than the coefficient on a younger sibling's height.[46] Table A.15 replicates the analysis in Table A.5 with heights instead of AFQT scores as explanatory variables. In the upper panel, where a middle sibling's AFQT score and schooling are regressed on her own height as well as those of her youngest and oldest siblings, the coefficient on an oldest sibling's height is insignificantly lower than the coefficient on a youngest sibling's height.[47] In the lower panel, where a youngest (resp. oldest) sibling's AFQT score and schooling are regressed on her own and her oldest (resp. youngest) sibling's heights, the coefficient on an oldest sibling's height is insignificantly less than the coefficient on a youngest sibling's height.[48] Overall, using height as an ability measure that is unlikely to be

[45] Information on height is available in the NLSY79 for the following survey years: 1982, 1983, 1985, 2006, and 2008. The most recent height observation on an individual is used for the analysis. Because some respondents may not have reached their adult height when surveyed, the estimation was also performed excluding respondents whose height was recorded only before reaching age 20. The findings are similar regardless of whether such individuals are dropped.

[46] With and without including family background controls, the respective two-sided p-values for the test of equality between these coefficients are 0.2196 and 0.2851 when the AFQT score is the dependent variable and 0.3273 and 0.3594 when schooling is the dependent variable.

[47] With and without including family background controls, the respective two-sided p-values for the test of equality between these coefficients are 0.5795 and 0.5618 when the AFQT score is the dependent variable and 0.6741 and 0.6317 when schooling is the dependent variable.

[48] With and without including family background controls, the respective two-sided p-values for the test of equality between these coefficients are 0.1214 and 0.1219 when the AFQT score is the dependent variable and 0.1519 and

affected by schooling, I continue to find no evidence that an older sibling has a stronger impact than a younger sibling on pre-labor market indicators of skills.[49]

A.21 AFQT Impacts on Joint Work-Wage Outcomes

As noted in section 1.5, a potential issue with the selection criteria used to generate the main estimation sample is that the analysis is restricted to sibling pairs in which both members have worked since the last interview. However, selection into employment is unlikely to be entirely exogenous. Therefore, this appendix reports the impacts of one's own and a sibling's AFQT scores on the probability of having worked since the last interview as well as the joint probability of having worked and earned a wage above a given level.[50] To perform this extension, the main estimation sample is expanded to include observations on sibling pairs in which one or both members may not have valid wage data on a full-time job. Using information on the number of weeks worked since the previous interview, I generate an indicator equal to one if the respondent worked since last being surveyed and equal to zero otherwise.[51] In addition, I generate two additional indicators of joint work-wage outcomes, the first equal to one if the respondent worked at an hourly wage of at least $\$5.00$ in 1982-1984 terms and equal to zero otherwise, the second equal to one if the respondent worked at an hourly wage of at least $\$10.00$ in 1982-1984 terms and equal to zero otherwise.[52] These indicator variables are then regressed on one's own and a sibling's AFQT scores and schooling as in the fifth and sixth columns of Table 1.3, controlling for the siblings'

0.1300 when schooling is the dependent variable.

[49]In addition, the estimates in Table 1.7 were reproduced using height instead of schooling as an explanatory variable. Consistent with one's own and a sibling's heights being easily observable to employers, there is no significant evidence of a difference between the impacts of one's own and an older sibling's heights on a younger sibling's log wage and the impacts of one's own and a younger sibling's heights on an older sibling's log wage.

[50]Other methods of accounting for the employment decision include the use of a median regression or a selection correction. However, such procedures are difficult to justify in the current setting because they usually rely on an assumption about the wage offers of nonparticipants relative to participants or the existence of a variable affecting participation but not wage offers.

[51]An observation on a sibling pair is excluded if one or both members are missing information on the number of weeks worked or have a positive number of weeks unaccounted for in the work history data.

[52]An observation on a sibling pair is eliminated from the sample if one or both members worked since the previous interview but lack valid wage information.

Table A.14: Relationship of Own AFQT and Schooling to Height of Younger or Older Sibling

	AFQT		Schooling	
Older Sibling's Height × Younger Sibling	0.0150 (0.0050)	0.0030 (0.0044)	0.0424 (0.0140)	0.0094 (0.0129)
Younger Sibling's Height × Older Sibling	0.0203 (0.0053)	0.0084 (0.0046)	0.0549 (0.0139)	0.0214 (0.0123)
Own Height	0.0281 (0.0048)	0.0142 (0.0043)	0.0772 (0.0129)	0.0386 (0.0119)
Family Background Controls	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.2851	0.2196	0.3594	0.3273
R^2	0.2749	0.4254	0.0852	0.2903
Families	1990	1990	1990	1990
Individuals	4720	4720	4720	4720
Sibling Pairs	7068	7068	7068	7068

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable and fixed effects for each of the two siblings' years of birth. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. All specifications include an indicator for whether the respondent is the older or the younger sibling in a given pair. The dataset used here contains the first observation on every sibling pair in the main estimation sample for which both members have information on height.

Table A.15: Relationship of Own AFQT and Schooling to Height of Youngest and Oldest Sibling

	Middle Siblings			
	AFQT		Schooling	
Oldest Sibling's Height	0.0006	-0.0021	0.0081	-0.0051
	(0.0137)	(0.0124)	(0.0385)	(0.0364)
Youngest Sibling's Height	0.0118	0.0071	0.0355	0.0173
	(0.0113)	(0.0097)	(0.0344)	(0.0303)
Own Height	0.0239	0.0086	0.0572	0.0197
	(0.0136)	(0.0132)	(0.0350)	(0.0336)
Family Background Controls	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.5618	0.5795	0.6317	0.6741
R^2	0.3510	0.5159	0.1723	0.3758
Families	496	496	496	496
Individuals	626	626	626	626
Sibling Trios	828	828	828	828
	Oldest and Youngest Siblings			
	AFQT		Schooling	
Oldest Sibling's Height $\times$ Youngest Sibling	0.0121	0.0009	0.0325	0.0015
	(0.0051)	(0.0046)	(0.0146)	(0.0136)
Youngest Sibling's Height $\times$ Oldest Sibling	0.0203	0.0083	0.0549	0.0207
	(0.0053)	(0.0047)	(0.0146)	(0.0128)
Own Height	0.0277	0.0146	0.0770	0.0399
	(0.0047)	(0.0043)	(0.0131)	(0.0120)
Family Background Controls	No	Yes	No	Yes
test for equality between sibling AFQT coefficients (p-value)	0.1219	0.1214	0.1300	0.1519
R^2	0.2754	0.4246	0.0873	0.2953
Families	1990	1990	1990	1990
Individuals	4678	4678	4678	4678
Sibling Pairs	6058	6058	6058	6058

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling trio or pair. Included also are indicators for missing data on a given variable and fixed effects for each of the siblings' years of birth. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the siblings' birth orders. In the lower panel, all specifications include an indicator for whether the respondent is the older or the younger sibling in a given pair. The dataset used in the upper panel contains the first observation on each sibling trio in the main estimation sample consisting of a middle sibling along with her youngest and her oldest sibling in the labor market, provided that all members of the sibling trio have information on height. The dataset used in the lower panel contains the first observation on every sibling pair in the main estimation sample composed of the oldest and the youngest sibling from a family in the labor market, provided that both members of the sibling pair have information on height.

demographic characteristics and computing separate estimates for younger and older siblings.

The results of this procedure are documented in Table A.16. The upper panel includes observations on sibling pairs in which one or both members may not yet have left school, and the lower panel is limited to observations on sibling pairs in which both members have left school for the first time.[53] In all cases, an older sibling's AFQT score has a greater estimated impact on a younger sibling's outcome than a younger sibling's AFQT score has on an older sibling's outcome, and the estimated impact of one's own AFQT score on one's own outcome is larger for an older than for a younger sibling. Testing whether the ratio of the coefficient on an older sibling's AFQT score to the coefficient on one's own AFQT score in a younger sibling's outcome equation is equal to the ratio of the coefficient on a younger sibling's AFQT score to the coefficient on one's own AFQT score in an older sibling's outcome equation, the restriction can be rejected when the dependent variable is an indicator for having worked at a wage of at least $\$5.00$.[54] When the dependent variable is an indicator for having worked since the last interview or an indicator for having worked at a wage of at least $\$10.00$, the ratio of the coefficient on an older sibling's AFQT score to the coefficient on one's own AFQT score in a younger sibling's outcome equation is higher than vice versa, although this difference is not statistically significant at the five percent level. In sum, after modifying the estimation procedure to account for non-working individuals, I continue to find some evidence that an older sibling has a bigger impact on a younger sibling's labor market outcomes than vice versa.

A.22 Instrumental-Variables Estimates of Schooling Coefficients

This appendix provides instrumental-variables estimates for the relationship of one's own and a sibling's schooling to one's test score and log wage. In Tables A.10 and A.12, I found that the coefficient on a sibling's schooling is significantly positive in the regression of one's test score or log wage on both one's own and a sibling's schooling. As explained in section 1.2.2, this finding may indicate that the sibling correlation in ability is greater than that in schooling. However,

[53] In the upper panel, the percentages of observations on respondents working since the last interview, working for a wage of at least $\$5.00$, and working for a wage of at least $\$10.00$ are respectively 81.82, 49.53, and 14.71. In the lower panel, these percentages are 84.00, 53.97, and 16.63, respectively.

[54] With and without including family background controls, the respective two-sided p-values for this test are 0.0476 and 0.0438 in the upper panel and 0.0384 and 0.0362 in the lower panel.

Table A.16: Impact of Own AFQT and AFQT of Younger or Older Sibling on Joint Work-Wage Outcomes

	Entire Sample					
	Worked		Wage ≥ \$5 Also		Wage ≥ \$10 Also	
Older Sibling's AFQT × Younger Sibling	0.0103 (0.0053)	0.0121 (0.0054)	0.0148 (0.0070)	0.0157 (0.0070)	0.0110 (0.0057)	0.0082 (0.0055)
Younger Sibling's AFQT × Older Sibling	0.0094 (0.0055)	0.0102 (0.0052)	-0.0033 (0.0074)	-0.0015 (0.0072)	-0.0023 (0.0059)	-0.0043 (0.0058)
Own AFQT × Younger Sibling	0.0260 (0.0053)	0.0279 (0.0054)	0.0615 (0.0074)	0.0624 (0.0074)	0.0393 (0.0056)	0.0366 (0.0057)
Own AFQT × Older Sibling	0.0356 (0.0061)	0.0350 (0.0062)	0.0993 (0.0084)	0.0982 (0.0083)	0.0653 (0.0072)	0.0626 (0.0073)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.6546	0.6194	0.0438	0.0476	0.0753	0.1061
R^2	0.1725	0.1836	0.2558	0.2632	0.2049	0.2090
Families	2181	2181	2181	2181	2181	2181
Individuals	5195	5195	5195	5195	5195	5195
Sibling Pairs	8032	8032	8032	8032	8032	8032
Observations	123388	123388	123388	123388	123388	123388
	Out of School					
	Worked		Wage ≥ \$5 Also		Wage ≥ \$10 Also	
Older Sibling's AFQT × Younger Sibling	0.0146 (0.0057)	0.0167 (0.0057)	0.0197 (0.0080)	0.0202 (0.0080)	0.0150 (0.0067)	0.0114 (0.0065)
Younger Sibling's AFQT × Older Sibling	0.0060 (0.0063)	0.0072 (0.0059)	-0.0018 (0.0086)	-0.0008 (0.0084)	0.0037 (0.0071)	0.0006 (0.0070)
Own AFQT × Younger Sibling	0.0285 (0.0059)	0.0306 (0.0060)	0.0794 (0.0088)	0.0798 (0.0088)	0.0544 (0.0070)	0.0506 (0.0071)
Own AFQT × Older Sibling	0.0374 (0.0067)	0.0374 (0.0067)	0.1134 (0.0093)	0.1120 (0.0092)	0.0780 (0.0082)	0.0745 (0.0082)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.2301	0.2142	0.0362	0.0384	0.1541	0.1878
R^2	0.1210	0.1342	0.2303	0.2388	0.2019	0.2165
Families	2161	2161	2161	2161	2161	2161
Individuals	5149	5149	5149	5149	5149	5149
Sibling Pairs	7948	7948	7948	7948	7948	7948
Observations	104602	104602	104602	104602	104602	104602

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's outcome is used as the dependent variable for a given pair. The datasets used here are constructed as follows. First, the main estimation sample is expanded to include observations on sibling pairs in which one or both members may not have valid wage data on a full-time job. Second, the resulting sample is restricted to observations on sibling pairs in which both members have non-missing data on the number of weeks worked since the last interview. An observation on a sibling pair is excluded from the analysis if one or both members worked since the last interview but do not have valid wage data. In the upper panel, the sample contains observations on sibling pairs in which one or both members may not yet have left school. In the lower panel, the sample includes only observations on sibling pairs in which both members have left school for the first time. In the first pair of columns, the dependent variable is an indicator equal to one if the respondent worked since the last interview and equal to zero otherwise. The dependent variable in the second (third) pair of columns is an indicator equal to one if the respondent worked since the last interview at an hourly wage of at least \$5.00 (\$10.00) in 1982-1984 terms and equal to zero otherwise.

the presence of measurement error in the schooling variable complicates the interpretation of the result. In this appendix, I attempt to correct for measurement error by using a variant of the instrumental-variables procedure in Ashenfelter and Krueger (1994) and Bronars and Oettinger (2006). Specifically, respondents in the NLSY79 were asked to report the highest grade completed by each of their siblings when interviewed in 1993, making it possible to use one sibling's reports of both siblings' schooling as an instrument for the other sibling's reports of these variables in that year.

In the current context, a potential issue with the instrumental-variables procedure is that one sibling's reports of both siblings' schooling will not be a valid instrument for the other sibling's reports of these variables if the measurement error in schooling is correlated between siblings. Nonetheless, there are two arguments for proceeding with the instrumental-variables strategy. First, the two-stage least squares estimator is valid if there is no correlation in measurement error between siblings. As shown in appendix A.9, this is precisely the case in which the interpretation of the positive coefficient on a sibling's schooling as indicating a greater sibling correlation in ability than in schooling is most likely to be incorrect when schooling is measured with a certain amount of classical error. Hence, the instrumental-variables procedure should be relatively accurate in this most critical case. Second, appendix A.11 examines the two-stage least squares regression of one's test score or log wage on both one's own and a sibling's schooling in which a sibling's reports of the latter two variables are used as instruments for one's own reports of these variables, even though there is a positive correlation between the errors in the reports of the two siblings. Under reasonable assumptions, I show that the parameter on a sibling's schooling identified by this regression is less than the coefficient on a sibling's schooling in the conditional expectation of one's test score or log wage given one's own and a sibling's true schooling. Therefore, if one continues to observe that the coefficient on a sibling's schooling is significantly positive when using the instrumental-variables estimator, then this finding can likely be regarded as evidence of a higher sibling correlation in ability than in schooling.

The instrumental-variables procedure is implemented as follows. First, I construct a dataset containing all sibling pairs belonging to the sample from Table A.11 in 1993 for which both members have non-missing data on sibling-reported schooling. This dataset consists of 2740 sibling pairs from 902 families. Second, I perform a two-stage least squares regression of the first sibling's test score on both her own and the second sibling's schooling, where the second sibling's reports of her own and the first sibling's schooling are used as instruments for the first sibling's reports of her own and the second sibling's schooling. The estimated specification controls for both one's own and a sibling's race, gender, region of residence, urban location, birth order, sibship size, and birth year. The Huber-White estimator of the variance-covariance matrix with clustering at the family level is used to account for the correlation in the error terms among individuals belonging to the same sibship.

In the first-stage regression of the first sibling's reports of her own and the second sibling's schooling on the second sibling's reports of her own and the first sibling's schooling as well as the other control variables, the point estimate (standard error) is 0.0541 (0.0144) for the coefficient on the second sibling's report of her own schooling and 0.9085 (0.0179) for the coefficient on the second sibling's report of the first sibling's schooling when the dependent variable is the first sibling's report of her own schooling, and the point estimate (standard error) is 0.7506 (0.0152) for the coefficient on the second sibling's report of her own schooling and 0.0930 (0.0156) for the coefficient on the second sibling's report of the first sibling's schooling when the dependent variable is the first sibling's report of the second sibling's schooling. The R-squared value is 0.7757 for the former and 0.7803 for the latter regression. The corresponding two-stage least squares estimates (standard errors) for the coefficients on one's own and a sibling's schooling are respectively 0.2127 (0.0098) and 0.0338 (0.0104) when the dependent variable is the AFQT score. Hence, I continue to observe a significantly positive coefficient on a sibling's schooling when regressing one's AFQT score on both one's own and a sibling's schooling. Using the AFQT score as an ability measure, the earlier finding in Table A.10 of a higher sibling correlation in ability than in schooling is not likely to be attributable merely to measurement error in the schooling variable.

I also report the results of the instrumental-variables procedure when the dependent variable is the log wage instead of the AFQT score. The dataset used here includes any sibling pair belonging to the sample in Table A.13 during the 1993 survey year, provided that both members of the pair have non-missing data on sibling-reported schooling. On this dataset composed of 1448 sibling pairs from 522 families, I run a two-stage least squares regression of the first sibling's log wage on both her own and the second sibling's schooling, where the second sibling's reports of the two siblings' schooling are used as instruments for the first sibling's reports of the two siblings' schooling. In addition, I control for both one's own and a sibling's race, gender, region of residence, urban location, birth order, and sibship size as well as a third-order bivariate polynomial in the ages of the siblings in each pair. As before, the standard errors are clustered at the family level.

In the first-stage regression of the first sibling's reports of the two siblings' schooling on the second sibling's reports of the two siblings' schooling along with the other control variables, the point estimate (standard error) is 0.0507 (0.0230) for the coefficient on the second sibling's report of her own schooling and 0.8942 (0.0253) for the coefficient on the second sibling's report of the first sibling's schooling when the dependent variable is the first sibling's report of her own schooling, and the point estimate (standard error) is 0.7636 (0.0215) for the coefficient on the second sibling's report of her own schooling and 0.0926 (0.0186) for the coefficient on the second sibling's report of the first sibling's schooling when the dependent variable is the first sibling's report of the second sibling's schooling. The R-squared value is 0.7744 for the former and 0.7798 for the latter regression. Using the log wage as the dependent variable, the two-stage least squares estimates (standard errors) for the coefficients on one's own and a sibling's schooling are respectively 0.0888 (0.0099) and 0.0188 (0.0094). As in Table A.12, the significantly positive coefficient on a sibling's schooling suggests that the sibling correlation in ability is higher than that in schooling.

Appendix B

Appendices to Chapter 3

B.1 Proof of Proposition 3.4.1

Proof Because transaction costs are paid only a countable number of times, s_t^{12} and s_t^{21} are step functions in time, which we define to be continuous from the right. Letting $t(k-1;\pi) = 0$ if $k = 1$, we have for $j \neq i$:

$$
\begin{aligned}
&\int_d^\infty e^{-\rho\cdot(t-d)} s_t^{ij} dt \;-\; \sum_{\{k\colon\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} C_{t(k;\pi)} \\
&= \int_d^\infty e^{-\rho\cdot(t-d)} s_{t(k-1;\pi)}^{ij} dt + \sum_{\{k\colon\, t(k;\pi)\geq d\}} \int_{t(k;\pi)}^\infty e^{-\rho\cdot(\tau-d)} (s_{t(k;\pi)}^{ij} - s_{t(k-1;\pi)}^{ij}) d\tau \\
&\quad - \sum_{\{k\colon\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} C_{t(k;\pi)} \\
&= s_{t(k-1;\pi)}^{ij} \cdot -\frac{1}{\rho} e^{-\rho\cdot(\tau-d)} \bigg|_d^\infty + \sum_{\{k\colon\, t(k;\pi)\geq d\}} (s_{t(k;\pi)}^{ij} - s_{t(k-1;\pi)}^{ij}) \cdot -\frac{1}{\rho} e^{-\rho\cdot(\tau-d)} \bigg|_{t(k;\pi)}^\infty \\
&\quad - \sum_{\{k\colon\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} C_{t(k;\pi)} \\
&= \frac{1}{\rho} s_{t(k-1;\pi)}^{ij} + \frac{1}{\rho} \sum_{\{k\colon\, t(k;\pi)\geq d\}} (s_{t(k;\pi)}^{ij} - s_{t(k-1;\pi)}^{ij}) e^{-\rho\cdot[t(k;\pi)-d]} \\
&\quad - \frac{1}{\rho} \sum_{\{k\colon\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} \rho C_{t(k;\pi)}
\end{aligned}
$$

$$= \frac{1}{\rho}\left(s^{ij}_{t(k-1;\pi)} + \sum_{\{k:\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} x^{j}_{t(k;\pi)} - \sum_{\{k:\, t(k;\pi)\geq d\}} e^{-\rho\cdot[t(k;\pi)-d]} c_{t(k;\pi)} \right),$$

where we set $c_{t(k;\pi)} = \rho C_{t(k;\pi)}$.

B.2 Proof of Theorem 3.4.2

Proof We prove by induction that no strategies in which some agent i makes a positive transfer x^i_t at some history h_t can be played in any SPE.

Consider any history h_t where $\hat{s}^{ii}_t \in [0, q]$ is the amount of agent i's good remaining up to but not including time t. If $\hat{s}^{11}_t < \underline{c}$, then no strategy where agent 2 makes a positive transfer at history h_t can be played in any SPE, because such a strategy would give agent 2 an expected discounted payoff no greater than $\hat{s}^{11}_t - \underline{c} < 0$, whereas agent 2 could obtain an expected discounted payoff of at least zero by making no transfers after history h_t. Because no strategy where agent 2 makes a positive transfer at some history h_t with $\hat{s}^{11}_t < \underline{c}$ can be played in any SPE, agent 1 would obtain an expected discounted payoff no greater than $-\underline{c} < 0$ by making a positive transfer at a history h_t with $\hat{s}^{11}_t < \underline{c}$ but would obtain an expected discounted payoff of zero by making no transfers after such a history. Thus, no strategy where agent 1 makes a positive transfer at some history h_t in which $\hat{s}^{11}_t < \underline{c}$ can be played in any SPE. A symmetric argument shows that no strategy where agent 1 or 2 makes a positive transfer at some history h_t in which $\hat{s}^{22}_t < \underline{c}$ can be played in any SPE.

Suppose now that for some integer $n \geq 1$, no strategies where agent 1 or 2 makes a positive transfer at some history h_t in which $\hat{s}^{11}_t < n\underline{c}$ or $\hat{s}^{22}_t < n\underline{c}$ can be played in any SPE. Given this assumption, we show that no strategies where agent 1 or 2 makes a positive transfer at some history h_t in which $n\underline{c} \leq \hat{s}^{11}_t < (n+1)\underline{c}$ or $n\underline{c} \leq \hat{s}^{22}_t < (n+1)\underline{c}$ can be played in any SPE. Consider in particular a history h_t in which $n\underline{c} \leq \hat{s}^{11}_t < (n+1)\underline{c}$. If agent 1 is using a strategy that can be played in an SPE, then the greatest amount of the good that agent 1 can transfer at history h_t is $\hat{s}^{11}_t - n\underline{c}$, because if agent 1 made a transfer greater than $\hat{s}^{11}_t - n\underline{c}$ at history h_t, then the remaining amount $\hat{s}^{11}_\tau$ of agent 1's good for $\tau > t$ would be such that agent 2 makes no further transfers, implying that agent 1 could obtain a higher expected discounted payoff by instead making no transfers after history h_t. Thus, if agent 1 is using a strategy that can be played in an SPE and history h_t is reached, then it must be that $\hat{s}^{11}_\tau \geq n\underline{c}$ for every history h_τ with $\tau > t$; so that, the total amount transferred by agent 1 after history h_t is at most $\hat{s}_t^{11} - n\underline{c} < \underline{c}$.

It follows that no strategy in which agent 2 makes a positive transfer after history h_t can be played in any SPE, because making a positive transfer would give agent 2 an expected discounted payoff no greater than $\hat{s}_t^{11} - n\underline{c} - \underline{c} < 0$, whereas agent 2 could obtain an expected discounted payoff of at least zero by making no transfers after history h_t. Thus, no strategy where agent 1 makes a transfer at history h_t can be played in any SPE, because agent 1 would obtain an expected discounted payoff no greater than $-\underline{c} < 0$ by making a transfer at history h_t but would obtain an expected discounted payoff of zero by making no transfers after history h_t. Thus, no strategies where agent 1 or 2 makes a transfer at some history h_t in which $n\underline{c} \leq \hat{s}_t^{11} < (n+1)\underline{c}$ can be played in any SPE. A symmetric argument holds for any history h_t in which $n\underline{c} \leq \hat{s}_t^{22} < (n+1)\underline{c}$. This completes the induction.

B.3 Proof of Proposition 3.4.3

Proof Fix any SPE $\pi = (\pi_1, \pi_2)$. We construct an SPE that induces the same equilibrium path of play using grim-trigger strategies. Given any history h_t, let $u_i(h_t, \pi)$ denote the expected discounted payoff to agent i when both agents follow the strategy profile π from time t onwards, and let $v_i(h_t, \pi)$ denote the supremum of the expected discounted payoffs to agent i from any set of deviations when π_{-i} is fixed.[1] Because π is an SPE, it must be that $u_i(h_t, \pi) \geq v_i(h_t, \pi)$.

Now consider the grim-trigger strategy profile $\pi' = (\pi'_1, \pi'_2)$ such that $\pi'(h_t) = \pi(h_t)$ for any history h_t on the equilibrium path induced by π and such that $\pi'(h_t) = 0$ for any history h_t not on the equilibrium path induced by π.[2] We show that π' is an SPE strategy profile.[3]

Suppose that the history h_t is on the equilibrium path induced by π. If $\pi'_i(h_t) = 0$, then the expected discounted payoff to agent i from a one-shot deviation at history h_t must be negative. However, the expected discounted payoff to agent i when both agents follow the strategy profile π' from history h_t onwards must be nonnegative, because $\pi'(h_t) = \pi(h_t)$ for any history h_t on the

[1]Note that the index $-i$ is used to denote the agent other than i.

[2]Recall that $\pi_i(h_t)$ is the amount that agent i transfers at time t conditional on history h_t when using strategy π_i.

[3]When confirming that the grim-trigger strategy profile π' is an SPE, it is sufficient to consider only one-shot deviations from π'.

equilibrium path induced by π, where π is an SPE. Hence, if $\pi'_i(h_t) = 0$, then agent i does not have an incentive to make a one-shot deviation at history h_t.

Assume instead that $\pi'_i(h_t) > 0$. By definition, if agent $-i$ is using strategy π'_{-i}, then the expected discounted payoff to agent i from following π'_i from history h_t onwards is $u_i(h_t, \pi)$; so that, $u_i(h_t, \pi') = u_i(h_t, \pi)$. In addition, if agent $-i$ is using strategy π'_{-i}, then a one-shot deviation at history h_t gives agent i an expected discounted payoff of at most $\pi'_{-i}(h_t)$, because agent $-i$ makes no transfers at any history not on the equilibrium path induced by π. Furthermore, agent i can ensure that she receives an expected discounted payoff of $\pi'_{-i}(h_t)$ by transferring nothing from time t onward. Hence, $\tilde{v}_i(h_t, \pi') = \pi'_{-i}(h_t)$, where $\tilde{v}_i(h_t, \pi')$ is the supremum of the expected discounted payoffs to agent i at history h_t from any one-shot deviation when π'_{-i} is fixed.

Now note that $v_i(h_t, \pi) \geq \pi_{-i}(h_t) = \pi'_{-i}(h_t)$ because if agent $-i$ is using strategy π_{-i}, then agent i can always obtain at history h_t an expected discounted payoff of at least $\pi_{-i}(h_t)$ by transferring nothing from time t onward. Thus, we have $u_i(h_t, \pi') = u_i(h_t, \pi) \geq v_i(h_t, \pi) \geq \pi'_{-i}(h_t) = \tilde{v}_i(h_t, \pi')$ for any history h_t on the equilibrium path induced by π such that $\pi'_i(h_t) > 0$. Moreover, for any history h_t not on the equilibrium path induced by π, we have $u_i(h_t, \pi') \geq v_i(h_t, \pi')$ because $u_i(h_t, \pi') = 0 = v_i(h_t, \pi')$. Hence, the strategy profile π' constitutes an SPE. Since π' is defined so as to agree with π on the equilibrium path, this completes the proof.

B.4 Proof of Corollaries to Theorem 3.4.4

B.4.1 Proof of Corollary 3.5.4

Proof The sign of $\partial x^*_k / \partial \beta$ can be determined as follows:

$$\begin{aligned}
\operatorname{sgn}\left[\frac{\partial x^*_k}{\partial \beta}\right] &= \operatorname{sgn}\left[\partial\left(\frac{q}{1-\beta}\left(\frac{\beta}{\beta-1}\right)^{k-1}\right) \Big/ \partial\beta\right] \\
&= \operatorname{sgn}\left[-(k-1)(\beta)^{k-2}(\beta-1)^{k} + k(\beta)^{k-1}(\beta-1)^{k-1}\right] \\
&= \operatorname{sgn}\left[[-(k-1)(\beta-1) + k\beta](\beta)^{k-2}(\beta-1)^{k-1}\right] \\
&= \operatorname{sgn}\left[(k-1)(\beta-1) - k\beta\right] = \operatorname{sgn}\left[1-k-\beta\right].
\end{aligned}$$

Thus, $\partial x_k^*/\partial\beta$ is negative if $k-1 > |\beta|$ and positive if $k-1 < |\beta|$. In addition, we have $\partial\beta/\partial\mu < 0$ and $\partial\beta/\partial(\sigma^2) > 0$. Combining these facts, we obtain the desired result.

B.4.2 Proof of Corollary 3.5.5

Proof The sign of $\partial c_k^*/\partial\beta$ can be determined as follows:

$$\begin{aligned}
\operatorname{sgn}\left[\frac{\partial c_k^*}{\partial\beta}\right] &= \operatorname{sgn}\left[\frac{\partial\ln(c_k^*)}{\partial\beta}\right] \\
&= \operatorname{sgn}\left[\frac{\partial\{\ln(q)-\ln(1-\beta)+(k-\beta)\left[\ln(-\beta)-\ln(1-\beta)\right]\}}{\partial\beta}\right] \\
&= \operatorname{sgn}\left[\frac{1}{1-\beta}+(k-\beta)\left(-\frac{1}{\beta}+\frac{1}{1-\beta}\right)-\left[\ln(-\beta)-\ln(1-\beta)\right]\right] \\
&= \operatorname{sgn}\left[\frac{1}{1-\beta}+(k-\beta)\left(\frac{1}{-\beta}+\frac{1}{1-\beta}\right)+\left[\ln(1-\beta)-\ln(-\beta)\right]\right].
\end{aligned}$$

Thus, $\partial c_k^*/\partial\beta$ is positive. In addition, we have $\partial\beta/\partial\mu < 0$ and $\partial\beta/\partial(\sigma^2) > 0$. Combining these facts, we obtain the desired result.

B.5 Proofs of Theorem 3.6.1 and Corollary 3.6.2

B.5.1 Proof of Theorem 3.6.1

Proof We begin by computing the solution to the second-best problem. From the proof of Theorem 3.4.4, the value function $V^{sb}[c_0, s_0^{ii}; \beta(\rho)]$ of each agent for the second-best problem is as follows:

$$V^{sb}[c_0, s_0^{ii}; \beta(\rho)] = \frac{f[\beta(\rho)]s_0^{ii}}{\{h[s_0^{ii}; \beta(\rho)]\}^{\beta(\rho)}}(c_0)^{\beta(\rho)},$$

where $f[\beta(\rho)]$ and $h[s_0^{ii}; \beta(\rho)]$ are defined as follows:

$$f[\beta(\rho)] = \frac{1}{1-\beta(\rho)} \quad \text{and} \quad h[s_0^{ii}; \beta(\rho)] = \frac{s_0^{ii}}{1-\beta(\rho)}\left(\frac{-\beta(\rho)}{1-\beta(\rho)}\right)^{1-\beta(\rho)},$$

and $\beta(\rho) < 0$ is given by:

$$\beta(\rho) = \tfrac{1}{2} - \tfrac{\mu}{\sigma^2} - \sqrt{\left(\tfrac{\mu}{\sigma^2} - \tfrac{1}{2}\right)^2 + 2\tfrac{\rho}{\sigma^2}}.$$

The notation here differs from that used in the proof of Theorem 3.4.4 in that β appears as an argument in the value function. In addition, $\beta(\rho)$ is used to denote the value of β given the discount rate ρ.

Substituting for $f[\beta(\rho)]$ and $h[s_0^{ii}; \beta(\rho)]$ in the expression for $V^{sb}[s_0^{ii}, c_0; \beta(\rho)]$ results in:

$$V^{sb}[c_0, s_0^{ii}; \beta(\rho)] = \{a[\beta(\rho)]\}^{1-\beta(\rho)}(s_0^{ii})^{1-\beta(\rho)}(c_0)^{\beta(\rho)},$$

where the constant $a[\beta(\rho)] \in (0, 1)$ is given by:

$$a[\beta(\rho)] = [-\beta(\rho)]^{-\beta(\rho)}[1 - \beta(\rho)]^{-[1-\beta(\rho)]}.$$

We next compute the solution to the first-best problem. In the absence of incentive constraints, the optimal policy requires the agents to make at most one transfer. Given that the agents use a stationary threshold policy, the value function $V^{fb}[c_0, s_0^{ii}; \beta(\rho)]$ of each agent for the first-best problem can be obtained by choosing $\hat{c} \in (0, s_0^{ii})$ so as to maximize the value $W^{fb}[c_0, s_0^{ii}, \hat{c}; \beta(\rho)]$ of an asset that pays $s_0^{ii} - \hat{c}$ when the cost first reaches $\hat{c}$ starting from c_0. From the earlier analysis, the value of such an asset is:

$$W^{fb}[c_0, s_0^{ii}, \hat{c}; \beta(\rho)] = (s_0^{ii} - \hat{c})(\hat{c}/c_0)^{-\beta(\rho)}.$$

Differentiating with respect to $\hat{c}$ yields the following first-order condition for $\hat{c}^*$:

$$-\beta(\rho)(s_0^{ii} - \hat{c}^*)c_0^{-1}(\hat{c}^*/c_0)^{-\beta(\rho)-1} - (\hat{c}^*/c_0)^{-\beta(\rho)} = 0 \;\Rightarrow\; \hat{c}^* = -\beta(\rho)[1 - \beta(\rho)]^{-1}s_0^{ii},$$

where the derivative is positive for $\hat{c} < \hat{c}^*$ and negative for $\hat{c} > \hat{c}^*$. Thus, the value function for the first-best problem is:

$$V^{fb}[c_0, s_0^{ii}; \beta(\rho)] = a[\beta(\rho)](s_0^{ii})^{1-\beta(\rho)}(c_0)^{\beta(\rho)}.$$

We now show that $V^{sb}[c_0, s_0^{ii}; \beta(\rho)]$ converges to $V^{fb}[c_0, s_0^{ii}; \beta(\rho)]$ as ρ approaches zero from the right if and only if the condition $\mu \leq \sigma^2/2$ is satisfied. The limiting value of $\beta(\rho)$ is as follows:

$$\lim_{\rho \to 0^+} \beta(\rho) = \begin{cases} 0, & \text{if } \mu \leq \sigma^2/2 \\ 1 - 2\mu/\sigma^2, & \text{if } \mu > \sigma^2/2 \end{cases}.$$

Note that $\lim_{z \to 0^-} a(z) = \lim_{z \to 0^-} (-z)^{-z} \cdot \lim_{z \to 0^-} (1-z)^{-(1-z)}$, where $\lim_{z \to 0^-} (1-z)^{-(1-z)}$ is clearly equal to one, and $\lim_{z \to 0^-} (-z)^{-z}$ is easily shown to be one by taking the logarithm and applying L'Hôpital's rule. Thus, the limiting value of $a[\beta(\rho)]$ is given by:

$$\lim_{\rho \to 0^+} a[\beta(\rho)] = \begin{cases} 1, & \text{if } \mu \leq \sigma^2/2 \\ a(1 - 2\mu/\sigma^2), & \text{if } \mu > \sigma^2/2 \end{cases}.$$

It follows that:

$$\lim_{\rho \to 0^+} V^{sb}[c_0, s_0^{ii}; \beta(\rho)] = \begin{cases} s_0^{ii}, & \text{if } \mu \leq \sigma^2/2 \\ [a(1 - 2\mu/\sigma^2)]^{2\mu/\sigma^2} (s_0^{ii})^{2\mu/\sigma^2} (c_0)^{1-2\mu/\sigma^2}, & \text{if } \mu > \sigma^2/2 \end{cases},$$

$$\lim_{\rho \to 0^+} V^{fb}[c_0, s_0^{ii}; \beta(\rho)] = \begin{cases} s_0^{ii}, & \text{if } \mu \leq \sigma^2/2 \\ a(1 - 2\mu/\sigma^2) (s_0^{ii})^{2\mu/\sigma^2} (c_0)^{1-2\mu/\sigma^2}, & \text{if } \mu > \sigma^2/2 \end{cases}.$$

Note that if $\mu > \sigma^2/2$, then:

$$[a(1 - 2\mu/\sigma^2)]^{2\mu/\sigma^2} < a(1 - 2\mu/\sigma^2);$$

so that, $V^{sb}[c_0, s_0^{ii}; \beta(\rho)]$ does not converge to $V^{fb}[c_0, s_0^{ii}; \beta(\rho)]$ in this case.

B.5.2 Proof of Corollary 3.6.2

Proof Let $\gamma = \mu/\sigma^2$. Then the ratio of $V^{sb}[c_0, s_0^{ii}; \beta(\rho)]$ to $V^{fb}[c_0, s_0^{ii}; \beta(\rho)]$ is given by:

$$[a(1 - 2\gamma)]^{-(1-2\gamma)}.$$

Recall that $a(\beta) = (-\beta)^{-\beta}(1-\beta)^{-(1-\beta)}$. Taking the logarithm of both sides, we obtain:

$$\ln[a(\beta)] = -\beta\ln(-\beta) - (1-\beta)\ln(-\beta).$$

Differentiating with respect to β, we have:

$$\frac{\partial \ln[a(\beta)]}{\partial\beta} = -\ln(-\beta) - 1 + \ln(-\beta) - \frac{(1-\beta)}{\beta} = -\frac{1}{\beta} > 0.$$

Hence, $a(\beta)$ is increasing in β. Consequently, both $(1-2\gamma)$ and $a(1-2\gamma)$ are decreasing in γ. Because $a(\beta) \in (0,1)$ and $(1-2\gamma) < 0$, the ratio is decreasing in γ. Since γ is increasing in μ and decreasing in σ^2, this completes the proof.

B.6 Proof of Theorem 3.7.1

Proof A symmetric SPE with a positive expected discounted payoff at each point along the path of play can be constructed as follows for the model with stochastic supplies of goods. Consider a symmetric strategy profile π such that each agent transfers a fraction $\tilde{f} \in (0,1)$ of the remaining stock of each good if the stock is currently $\tilde{z} > \chi$. Otherwise, if the stock is not currently $\tilde{z}$, then no transfer is made. Let $J(s_t^{ii})$ denote the expected discounted payoff to each agent under strategy profile π when the remaining stock of each good is currently $s_t^{ii} < \tilde{z}$. Then $J(s)$ is the same as the value of an asset that pays $\tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]$ at the first time that the stock reaches $\tilde{z}$ starting from s.

The Bellman equation for this asset-pricing problem is given by the following for $s \leq \tilde{z}$:

$$\rho J(s)dt = \mathbb{E}(dJ) \tag{B.1}$$

subject to the boundary condition $J(\tilde{z}) = \tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]$. A straightforward application of Ito's lemma to the preceding equation yields

$$\rho J(s) = \theta s\frac{\partial J(s)}{\partial s} + \frac{1}{2}\xi^2 s^2 \frac{\partial^2 J(s)}{\partial s^2},$$

which provides a second-order linear differential equation for $J(s)$. Seeking a solution of the form

$g(s;\tilde{f},\tilde{z}) = B(\tilde{f},\tilde{z})s^{\tilde{\beta}}$, the following quadratic equation is obtained by substituting the functional form into the differential equation

$$\tfrac{1}{2}\xi^2\tilde{\beta}(\tilde{\beta}-1) + \theta\tilde{\beta} - \rho = 0,$$

whose solution is given by

$$\tilde{\beta} = \tfrac{1}{2} - \tfrac{\theta}{\xi^2} \pm \sqrt{\left(\tfrac{\theta}{\xi^2} - \tfrac{1}{2}\right)^2 + 2\tfrac{\rho}{\xi^2}}.$$

Letting $\tilde{\beta}^{+}$ and $\tilde{\beta}^{-}$ respectively denote the positive and negative roots of the quadratic, the general solution to the differential equation is $J(s) = B^{+}(\tilde{f},\tilde{z})s^{\tilde{\beta}^{+}} + B^{-}(\tilde{f},\tilde{z})s^{\tilde{\beta}^{-}}$. It must be the case that $B^{-}(\tilde{f},\tilde{z}) = 0$, because $J(s)$ would otherwise become unboundedly large in absolute value as s goes to 0. Moreover, the boundary condition $J(\tilde{z}) = \tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]$ yields $B^{+}(\tilde{f},\tilde{z}) = \{\tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]\}/\tilde{z}^{\tilde{\beta}^{+}}$. Hence, the solution to the Bellman equation is

$$J(s) = \{\tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]\}(s/\tilde{z})^{\lambda},$$

where $\lambda = \tilde{\beta}^{+}$. Note that $\lambda > 1$ if $\theta < \rho$.

In order to construct an SPE with a positive expected discounted payoff at each point along the path of play, one can choose $\tilde{f}$, $\tilde{z}$ in the definition of π so as to maximize $J(s)$ subject to the constraint $J[(1-\tilde{f})\tilde{z}] = \chi$. This constrained maximization problem can be expressed as

$$\max_{\tilde{f},\tilde{z}} \{\tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]\}(s/\tilde{z})^{\lambda} \text{ s.t. } J[(1-\tilde{f})\tilde{z}] = \{\tilde{f}\tilde{z} - \chi + J[(1-\tilde{f})\tilde{z}]\}(1-\tilde{f})^{\lambda} = \chi,$$

which can be rewritten as

$$\max_{\tilde{f},\tilde{z}}\ \tilde{f}\tilde{z}(s/\tilde{z})^{\lambda} \text{ s.t. } \tilde{f}\tilde{z}(1-\tilde{f})^{\lambda} = \chi \quad\text{or}\quad \max_{\tilde{f}}\ \chi^{1-\lambda}s^{\lambda}\tilde{f}^{\lambda}(1-\tilde{f})^{\lambda^2-\lambda},$$

where the second expression follows from substituting $\tilde{z} = \tilde{f}^{-1}(1-\tilde{f})^{-\lambda}\chi$ into the maximand.

Taking the log of the maximand and eliminating constants, we obtain

$$\max_{\tilde{f}} \ \lambda \ln(\tilde{f}) + (\lambda^2 - \lambda) \ln(1 - \tilde{f}),$$

which is a concave function of $\tilde{f}$. The first-order condition is

$$\lambda/\tilde{f} - (\lambda^2 - \lambda)/(1 - \tilde{f}) = 0.$$

Hence, we have $\tilde{f} = \lambda^{-1}$. It follows that $\tilde{z} = \lambda[\lambda/(\lambda - 1)]^{\lambda}\chi$.

B.7 Proof of Proposition 3.7.2

Proof A symmetric SPE with a positive expected discounted payoff at each point along the path of play can be constructed as follows for the model with a cost proportional to the amount transferred, provided that the component of the cost proportional to the amount transferred is sufficiently small. Let $\epsilon \in (0, 1)$. Consider the symmetric grim-trigger strategy profile π_ϵ in which the k^{th} transaction is made when the fixed cost reaches $c_k^{**}(\epsilon) = (1-\epsilon)c_k^*$ for the first time, and the amount $x_k^{**}(\epsilon) = x_k^*$ is transferred by each agent at this transaction, where c_k^* and x_k^* are as defined in Theorem 3.4.4. Let $H(c, s; \epsilon)$ denote the expected discounted payoff under strategy profile π_ϵ in the model with a cost proportional to the amount transferred if the size of the fixed cost is currently c and the remaining stock of each good is s.

Note that for any positive integer k, the expected discounted payoff $H[c_k^{**}(\epsilon), q - \sum_{l=1}^{k} x_l^{**}(\epsilon); \epsilon]$ immediately after transaction k is increasing in ϵ for the model with a cost proportional to the amount transferred. This property holds because for any transaction $m > k$, the amount transferred at transaction m as well as the distribution of waiting times between transactions k and m do not change with ϵ, whereas the total cost paid at transaction m is decreasing in ϵ.

In addition, observe that $H[c_k^{**}(\phi), q - \sum_{l=1}^{k} x_l^{**}(\phi); \phi] = (1 - \phi)V(c_k^*, q - \sum_{l=1}^{k} x_l^*)$ for each transaction k, where $V(c_k^*, q - \sum_{l=1}^{k} x_l^*)$ is the expected discounted payoff to each agent immediately after transaction k in the basic model. This observation follows from the following facts. First, the value function $V(s, c)$ for the basic model is homogeneous of degree one. Second, if strategy profile π_ϕ is played in the model with a cost proportional to the amount transferred, then the fixed cost paid at each transaction k is a fraction $(1-\phi)$ of the total cost incurred at transaction k under the path of play in Theorem 3.4.4 for the basic model, and the amount transferred minus the component of the cost proportional to the amount transferred at each transaction k is a fraction $(1-\phi)$ of the amount transferred at transaction k under the path of play in Theorem 3.4.4 for the basic model. Third, if strategy profile π_ϕ is played in the model with a cost proportional to the amount transferred, then for any transaction $m > k$, the distribution of waiting times between transactions k and m is the same as it is under the path of play in Theorem 3.4.4 for the basic model.

Hence, the incentive constraint $c_k^{**}(\epsilon)+\phi\cdot x_k^{**}(\epsilon) \leq H[c_k^{**}(\epsilon), q-\sum_{l=1}^{k} x_l^{**}(\epsilon); \epsilon]$ will be satisfied at each transaction k when strategy profile π_ϵ is played in the model with a cost proportional to the amount transferred if the following condition holds for some $\epsilon > \phi$:

$$(1-\epsilon)c_k^* + \phi \cdot x_k^* \leq (1-\phi)V(c_k^*, q - \Sigma_{l=1}^{k} x_l^*).$$

Substituting $x_k^* = [\beta/(\beta-1)]^{\beta-1}c_k^*$ and $(1-\phi)V(c_k^*, q-\sum_{l=1}^{k} x_l^*) = (1-\phi)c_k^*$ into the preceding expression, we obtain

$$(1-\epsilon) + \phi \cdot [\beta/(\beta-1)]^{\beta-1} \leq (1-\phi)$$

after canceling out c_k^* on both sides. The inequality above is satisfied if $\epsilon \geq \phi\cdot\{1+[\beta/(\beta-1)]^{\beta-1}\}$. If $\phi < \{1+[\beta/(\beta-1)]^{\beta-1}\}^{-1}$, then there exists $\epsilon < 1$ satisfying the inequality above. Defining $\bar{\phi} = \{1+[\beta/(\beta-1)]^{\beta-1}\}^{-1}$, it follows that if $\phi < \bar{\phi}$, then there exists $\epsilon < 1$ for which π_ϵ is a symmetric SPE of the model with a cost proportional to the amount transferred such that each agent receives a positive expected discounted payoff at each point along the path of play.

References

Admati, Anat R. and Motty Perry, "Joint Projects without Commitment," *The Review of Economic Studies*, 1991, *58* (2), 259–276.

Almond, Douglas and Janet Currie, "Human Capital Development Before Age Five," in Orley Ashenfelter and David Card, eds., *The Handbook of Labor Economics*, Vol. 4b, Elsevier B.V., 2011, pp. 1315–1486.

Altonji, Joseph G. and Charles R. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo, "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Microeconomics*, 2010, *2* (4), 76–104.

Ashenfelter, Orley and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review*, 1994, *84* (5), 1157–1173.

Barron, John M., Mark C. Berger, and Dan A. Black, "Selective Counteroffers," *Journal of Labor Economics*, 2006, *24* (3), 385–409.

Beaudry, Paul and John DiNardo, "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 1991, *99* (4), 665–688.

Becker, Gary S., *The Economics of Discrimination*, 2nd ed., University of Chicago Press, 1971.

Behrman, Jere R. and Paul Taubman, "Birth Order, Schooling, and Earnings," *Journal of Labor Economics*, 1986, *4* (3), 121–145.

Bertola, Giuseppe and Ricardo Caballero, "Irreversibility and Aggregate Investment," *The Review of Economic Studies*, 1994, *61* (2), 223–246.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Black, Sandra E., Paul Devereux, and Kjell G. Salvanes, "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education," *Quarterly Journal of Economics*, 2005, *120* (2), 669–700.

Bradley, Robert H. and Bettye M. Caldwell, "The Relation of Home Environment, Cognitive Competence, and IQ among Males and Females," *Child Development*, 1980, *51* (4), 1140–1148.

_ , _ , and Stephen L. Rock, "Home Environment and School Performance: A Ten-Year Follow-Up and Examination of Three Models of Environmental Action," *Child Development*, 1988, *59* (4), 852–867.

Bronars, Stephen G. and Gerald S. Oettinger, "Estimates of the Return to Schooling and Ability: Evidence from Sibling Data," *Labour Economics*, 2006, *13* (1), 19–34.

Butcher, Kristin F. and Anne Case, "The Effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, 1994, *109* (3), 531–563.

Caldwell, Bettye M. and Robert H. Bradley, "Home Observation for Measurement of the Environment: Administration Manual," Technical Report, Family and Human Dynamics Research Institute, Arizona State University 2003.

Case, Anne and Christina Paxson, "Height, Health, and Cognitive Function at Older Ages," *American Economic Review: Papers and Proceedings*, 2008, *98* (2), 463–467.

Case, Anne C. and Lawrence F. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth," Working Paper, Harvard University 1991.

Coen, Robert M., "Labor Force and Unemployment in the 1920's and 1930's: A Reexamination Based on Postwar Experience," *Review of Economics and Statistics*, 1973, *55* (1), 46–55.

Compte, Olivier and Philippe Jehiel, "On the Role of Arbitration in Negotiations," Mimeo, C.E.R.A.S.-E.N.P.C. 1995.

_ **and** _ , "Voluntary Contributions to a Joint Project with Asymmetric Agents," *The Journal of Economic Theory*, 2003, *112* (2), 334–342.

_ **and** _ , "Gradualism in Bargaining and Contribution Games," *The Review of Economic Studies*, 2004, *71* (4), 975–1000.

Cunha, Flavio and James Heckman, "The Technology of Skill Formation," *American Economic Review: Papers and Proceedings*, 2007, *97* (2), 31–47.

DeGroot, Morris H., *Optimal Statistical Decisions*, McGraw-Hill, 1970.

Dehejia, Rajeev and Adriana Lleras-Muney, "Booms, Busts, and Babies' Health," *Quarterly Journal of Economics*, 2004, *119* (3), 1091–1130.

Dixit, Avinash, "Irreversible Investment with Price Ceilings," *The Journal of Political Economy*, 1991, *99* (3), 541–557.

Dutta, Prajit K., "A Folk Theorem for Stochastic Games," *The Journal of Economic Theory*, 1995, *66* (1), 1–32.

Elardo, Richard, Robert H. Bradley, and Bettye M. Caldwell, "A Longitudinal Study of the Relation of Infants' Home Environments to Language Development at Age Three," *Child Development*, 1977, *48* (2), 595–603.

Farber, Henry S. and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Fudenberg, Drew and Eric Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 1986, *54* (3), 533–554.

_ **and Yuichi Yamamoto**, "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," *The Journal of Economic Theory*, 2011, *146* (4), 1664–1683.

Golan, Limor, "Wage Signaling: A Dynamic Model of Intrafirm Bargaining and Asymmetric Learning," *International Economic Review*, 2009, *50* (3), 831–854.

Granovetter, Mark S., *Getting a Job: A Study of Contacts and Careers*, Harvard University Press, 1974.

Griliches, Zvi, "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 1977, *45* (1), 1–22.

_ , "Sibling Models and Data in Economics: Beginnings of a Survey," *Journal of Political Economy*, 1979, *87* (5), S37–S64.

Hellerstein, Judith K. and Melinda S. Morrill, "Dads and Daughters: The Changing Impact of Fathers on Women's Occupational Choices," *Journal of Human Resources*, 2011, *46* (2), 333–372.

Horner, Johannes, Takuo Sugaya, Satoru Takahashi, and Nicolas Vielle, "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \rightarrow 1$ and a Folk Theorem," *Econometrica*, 2011, *79* (4), 1277–1318.

Ioannides, Yannis M. and Linda D. Loury, "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, 2004, *42* (4), 1056–1093.

Kahn, Lisa B., "Asymmetric Information Between Employers," Working Paper, Yale School of Management 2009.

Kamada, Yuichiro and Michihiro Kandori, "Revision Games," Mimeo, Harvard University 2009.

Katz, Lawrence F. and Kevin M. Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Kessler, Daniel, "Birth Order, Family Size, and Achievement: Family Structure and Wage Determination," *Journal of Labor Economics*, 1991, *9* (4), 413–426.

Lam, David and Robert F. Schoeni, "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil," *Journal of Political Economy*, 1993, *101* (4), 710–740.

Lange, Fabian, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

Malmendier, Ulrike and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?," *Quarterly Journal of Economics*, 2011, *126* (1), 373–416.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

Marx, Leslie M. and Steven A. Matthews, "Dynamic Voluntary Contribution to a Public Project," *The Review of Economic Studies*, 2000, *67* (2), 327–358.

McDonald, Robert and Daniel Siegel, "The Value of Waiting to Invest," *The Quarterly Journal of Economics*, 1986, *101* (4), 707–727.

Montgomery, James D., "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *American Economic Review*, 1991, *81* (5), 1408–1418.

Murphy, Kevin M. and Finis Welch, "The Structure of Wages," *Quarterly Journal of Economics*, 1992, *107* (1), 285–326.

Neal, Derek A. and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 1996, *104* (5), 869–895.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates," *American Economic Journal: Applied Economics*, 2012, *4* (1), 1–29.

Pinkston, Joshua, "A Model of Asymmetric Employer Learning with Testable Implications," *Review of Economic Studies*, 2009, *76* (1), 367–394.

Pitchford, Rohan and Christopher M. Snyder, "A Solution to the Hold-Up Problem Involving Gradual Investment," *The Journal of Economic Theory*, 2004, *114* (1), 88–103.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Rees, Albert, "Information Networks in Labor Markets," *American Economic Review*, 1966, *56* (1/2), 559–566.

Rees, Albert E. and Wayne B. Gray, "Family Effects in Youth Employment," in Richard B. Freeman and David A. Wise, eds., *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, University of Chicago Press, 1982.

Romer, Christina, "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, 1986, *94* (1), 1–37.

Ruhm, Christopher J., "Are Recessions Good for Your Health?," *Quarterly Journal of Economics*, 2000, *115* (2), 617–650.

Simon, Leo K. and Maxwell B. Stinchcombe, "Extensive Form Games in Continuous Time: Pure Strategies," *Econometrica*, 1989, *57* (5), 1171–1214.

Solon, Gary, Mary Corcoran, Roger Gordon, and Deborah Laren, "A Longitudinal Analysis of Sibling Correlations in Economic Status," *Journal of Human Resources*, 1991, *26* (3), 509–534.

van den Berg, Gerard J., Maarten Lindeboom, and France Portrait, "Economic Conditions Early in Life and Individual Mortality," *American Economic Review*, 2006, *96* (1), 290–302.

Waldman, Michael, "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics*, 1984, *15* (2), 255–267.

Wooldridge, Jeffrey M., "Estimating Average Partial Effects Under Conditional Moment Independence Assumptions," Working Paper, Michigan State University 2004.

Zajonc, R. B., "Family Configuration and Intelligence," *Science*, 1976, *192* (4236), 227–236.